

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 6, 2004, 20:15:58 ; Search time 130 Seconds  
(Without alignments)  
804.498 Million cell updates/sec

Title: US-09-938-901a-2

Perfect score: 1701  
Sequence: 1 MEMARKALLAWYRENNARPLP.....VLRKALLPLAHAGVLPDPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodaca/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodaca/2/pubpaa/PTC\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodaca/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodaca/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodaca/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodaca/2/pubpaa/PTCUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodaca/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodaca/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodaca/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodaca/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodaca/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodaca/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodaca/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodaca/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodaca/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodaca/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodaca/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodaca/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1697	99.8	325 10 US-09-938-901-2	Sequence 2, Appl
2	483.5	28.4	486 9 US-09-925-301-1326	Sequence 1326, Ap
3	477	28.0	375 15 US-10-289-762-421	Sequence 421, App
4	448	26.3	482 12 US-10-425-114-68280	Sequence 68280, A
5	442	26.0	474 16 US-10-437-963-13871	Sequence 13871,
6	389.5	22.9	188 15 US-10-629-951-36	Sequence 36, Appl
7	377	22.2	313 14 US-10-156-761-12241	Sequence 12241, A
8	365	21.8	293 9 US-09-738-626-6433	Sequence 6433, Ap
9	337	19.5	1719 16 US-10-437-963-103808	Sequence 103808,
10	263.5	15.5	184 12 US-10-424-599-248081	Sequence 248081,
11	240.5	14.1	188 15 US-10-629-951-34	Sequence 34, Appl
12	234	13.8	230 12 US-10-335-977-7283	Sequence 7283, Ap
13	228.5	13.4	187 12 US-10-335-977-7282	Sequence 7282, Ap
14	200.5	11.8	98 12 US-10-335-977-7279	Sequence 7279, Ap
15	200.5	11.8	98 12 US-10-335-977-7280	Sequence 7280, Ap

16	200.5	11.8	214 12 US-10-335-977-7461	Sequence 7461, Ap
17	199	11.7	268 9 US-09-864-866-43	Sequence 43, Appl
18	191	11.2	185 15 US-10-629-951-35	Sequence 35, Appl
19	184.5	10.8	310 14 US-10-156-761-12127	Sequence 12127, A
20	176.5	10.4	90 11 US-09-864-408A-9042	Sequence 9042, Ap
21	176	10.3	259 15 US-10-369-493-5691	Sequence 5691, Ap
22	163.5	9.6	304 12 US-10-424-599-280749	Sequence 280749,
23	158.5	9.3	355 15 US-10-369-493-2113	Sequence 2113, Ap
24	155	9.1	172 12 US-10-424-599-150617	Sequence 150617,
25	155	9.1	393 16 US-10-437-963-113667	Sequence 113667,
26	150	8.8	373 12 US-10-425-114-65596	Sequence 65596, A
27	147.5	8.7	260 9 US-09-738-626-3828	Sequence 3828, Ap
28	145	8.5	211 9 US-09-912-020-296	Sequence 296, Appl
29	137.5	8.1	184 15 US-10-629-951-33	Sequence 33, Appl
30	130	7.6	158 12 US-10-424-599-246799	Sequence 246799,
31	129.5	7.6	170 16 US-10-767-701-39640	Sequence 39640, A
32	129.5	7.6	5245 14 US-10-329-079-11	Sequence 11, Appl
33	128	7.5	281 14 US-10-128-714-3128	Sequence 3128, Ap
34	128	7.5	461 14 US-10-128-714-8128	Sequence 8128, Ap
35	126	7.4	272 15 US-10-369-493-13299	Sequence 13299, A
36	125.5	7.4	148 16 US-10-767-701-61161	Sequence 61161, A
37	125.5	7.4	6238 16 US-10-343-710-71	Sequence 71, Appl
38	120.5	7.1	776 12 US-10-425-114-70995	Sequence 70995, A
39	119	7.0	187 12 US-10-425-114-68053	Sequence 68053, A
40	118.5	7.0	416 12 US-10-425-114-43512	Sequence 43512, A
41	117.5	6.9	1072 16 US-10-437-963-114893	Sequence 114893,
42	112.5	6.6	440 12 US-10-425-114-71777	Sequence 71777, A
43	112.5	6.6	1332 10 US-09-840-743-11	Sequence 11, Appl
44	111	6.5	1729 10 US-09-840-743-2	Sequence 2, Appl
45	110.5	6.5	457 12 US-10-282-122A-44928	Sequence 44928, A

## ALIGNMENTS

RESULT 1  
US-09-938-901-2  
Sequence 2, Application US/09938901  
Publication No. US20030008291A1  
GENERAL INFORMATION:  
APPLICANT: Kuramitsu Seiki,  
APPLICANT: Yokoyama Shigeaki  
TITLE OR INVENTION: GENE ENCODING DNA REPAIR ENZYME  
FILE REFERENCE: PH-1261-US  
CURRENT APPLICATION NUMBER: US/09/938, 901  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: JP2001-47762  
PRIOR FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 325  
TYPE: PRT  
ORGANISM: Thermus thermophilus  
US-09-938-901-2

Qy	1	MEMARKALLAWYRENNARPLPWRGKDPYRLVSEVLLQOTRVBOALPYRRFLERFTLK 60	99.8%; Score 1697; DB 10; Length 325;
Qy	1	MEMARKALLAWYRENNARPLPWRGKDPYRLVSEVLLQOTRVBOALPYRRFLERFTLK 60	Best Local Similarity 99.7%; Pred. No. 2e-151;
Db	1	VEAMRKALLAWYRENNARPLPWRGKDPYRLVSEVLLQOTRVBOALPYRRFLERFTLK 60	Matches 324; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy	61	ALAAASLEEVLRWQAGAGYRRAEHLRLARSEBELPPSFAELRGLPGLAPYTAASVSI 120	
Qy	61	ALAAASLEEVLRWQAGAGYRRAEHLRLARSEBELPPSFAELRGLPGLAPYTAASVSI 120	
Db	61	ALAAASLEEVLRWQAGAGYRRAEHLRLARSEBELPPSFAELRGLPGLAPYTAASVSI 120	
Qy	121	AFGERVAADVGNVRVSRLLPARESPYKELFALAOGLLEGVDPGVWQNALMELGATVC 180	
Qy	121	AFGERVAADVGNVRVSRLLPARESPYKELFALAOGLLEGVDPGVWQNALMELGATVC 180	
Db	121	AFGERVAADVGNVRVSRLLPARESPYKELFALAOGLLEGVDPGVWQNALMELGATVC 180	
Qy	181	LPRPRGACPLGAFRCGKAPGRYPAPRRBRAKEETLVALLGRKGVHLERLEGRFOG 240	
Qy	181	LPRPRGACPLGAFRCGKAPGRYPAPRRBRAKEETLVALLGRKGVHLERLEGRFOG 240	
Db	181	LPRPRGACPLGAFRCGKAPGRYPAPRRBRAKEETLVALLGRKGVHLERLEGRFOG 240	

```

Db      181 LPRKPRGACPLGAFGRGKAPRYPAPRRKAKERLVALVLLGKGVHLERLGRFGQ 240
Qy      241 LVGVPLPPEPELPGREAAFGVRSRPLGEVHALTHRLRLEVRGALMEGEGEDPMKRPPLP 300
Db      241 LVGVPLPPEPELPGREAAFGVRSRPLGEVHALTHRLRLEVRGALMEGEGEDPMKRPPLP 300
Qy      301 KMEKYLKRLPLLAAGVPLPDA 325
Db      301 KMEKYLKRLPLLAAGVPLPDA 325

```

# RESULT 2

```

US-09-925-301-1326
/ Sequence 1326, Application US/09925301
/ Patent No. US20020052308A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
/ FILE REFERENCE: PA106
/ CURRENT APPLICATION NUMBER: US/09/925,301
/ PRIOR FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: PCT/US00/05882
/ PRIOR FILING DATE: 2000-03-08
/ PRIOR APPLICATION NUMBER: 60/124,270
/ NUMBER OF SEQ ID NOS: 1694
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1326
/ LENGTH: 486
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (34)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (438)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (447)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1326

```

Query Match 28.4%; Score 483.5; DB 9; Length 486;  
 Best Local Similarity 36.5%; Pred. No. 1e-36;  
 Matches 135; Conservative 46; Mismatches 106; Indels 83; Gaps 13;

```

Qy      3 AMRKALLAWYRENAARPLPMRG-----EKDPYLVSEVLLQOTRVQALPYRRFLP 54
Db      106 AFRGSLISWYDQEKRLPMRRRAEDMDLDRAYAVVWSEVWLQOTQVATVINYYTGMWQ 165
Qy      55 RPFLLKALAAASLEEVLRWQAGYVRAEHLRLARS-VEEL---PSPFAEL--RGLPG 108
Db      166 KWPFLDPLASASLEEVNQWLAGGYSRGRRLQEGARKVVEELGGMPTAETLLOQLPG 225
Qy      109 LGYTLAAVAVSIFGERVAAVGVNRYLSRLPA-----RSPKEXELFALAQGLPEGV 163
Db      226 VGYTTGALASINFGATGVVDGNVAVLCRVNAGADPSSSTLVSQLMGLAQQLV-DPA 284
Qy      164 DPGVNNQALMELGATVCLPRPRPGACPLGAFGRKE-----A 201
Db      285 RPDGFWQAMELGATVCTPRPLCSGCPRESICRARRQREVSQALLASGLSSPVEECA 344
Qy      202 P-----GRYPAPRRKRAK--EERLVALVL-----LGRKGVHLER- 233
Db      345 PVTGQCHCLPPESEPDQTLGVVNFPRKRSRKPRESSATVLEQPGALGQILLVGRP 404
Qy      234 LGRFGQGLYGV--LPPPEELPGREAF-----GVRSPRLGEVHALTHRLRV 280
Db      405 NSGLLAGLMEFBSYVWSEBQQLKALLQELQKXAGPLPATHTXRLHGEVHTFSHIKLTY 464
Qy      281 EVRGALMEGE 290

```

```

Db      465 QVYGLALMEGQ 474

```

# RESULT 3

```

US-10-289-762-421
/ Sequence 421, Application US/10289762
/ Publication No. US20040006218A1
/ GENERAL INFORMATION:
/ APPLICANT: Griffiths, R.
/ TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
/ FILE REFERENCE: 9710-003-999
/ CURRENT APPLICATION NUMBER: US/10/289,762
/ PRIOR FILING DATE: 2003-03-27
/ NUMBER OF SEQ ID NOS: 6849
/ SEQ ID NO 421
/ LENGTH: 375
/ TYPE: PRT
/ ORGANISM: Chlamydia pneumoniae
US-10-289-762-421

```

Query Match 28.0%; Score 477; DB 15; Length 375;  
 Best Local Similarity 38.6%; Pred. No. 3e-36;  
 Matches 117; Conservative 44; Mismatches 112; Indels 30; Gaps 7;

```

Qy      6 KALLAWYRENAARPLPMRGKDPYVLYSEVLLQOTRVQALPYRRFLERFPTLKALAAA 65
Db      22 EALKKPFKRNKRSLLPMRNPTPTYSVWSEVWLQOTRAVVIDYFNQWMEFRPTIESLAAA 81
Qy      66 SLEEVLRWQAGYVRAEHLRLARSVEE---LPPSFAELRGLPGLPYTAAVASI 120
Db      82 KEEDVTKLMEGLGYSSARHLLEGARVWMEFHGKIPDDAISLAQRGVGPYVHALIAR 141
Qy      121 AFGERVAVDGNVNRVLSRLPARESPEKE-----LFLAQGLLPGCVDDGVNNQALMEL 175
Db      142 AFRRAAAVDGNVNRVLSRLPLETSTIDLESTRTWVSRIQALLPH-KSPREVIAEALIEL 200
Qy      176 GATVCLPFRPRGACPLGAFGRGKAPRYPAPRRKRAK-----ERLVALVL--GRKGVN 230
Db      201 GACIC-KVFPQCHRCFVRQACGAMRENKQFVLVRRARKKVIYTLHLVALVLYDGLVVE 259
Qy      231 LERLEGRFQGLYGVPLPPEPELPGREAAFGVRSR-----PLGEVHALTHRR 277
Db      260 KRRFKEMAGLYEPPYIEVSEBGLQDIEGFTYKMELESPLFELGNKEQGHAFTHNK 319
Qy      278 LRV 280
Db      320 VHL 322

```

# RESULT 4

```

US-10-425-114-68280
/ Sequence 68280, Application US/10425114
/ Publication No. US20040034888A1
/ GENERAL INFORMATION:
/ APPLICANT: Liu, Jindong
/ APPLICANT: Zhou, Yina
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Screen, Steven E.
/ APPLICANT: Tabaska, Jack E.
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(5313)B
/ CURRENT APPLICATION NUMBER: US/10/425,114
/ PRIOR FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 73128
/ SEQ ID NO 68280
/ LENGTH: 482
/ TYPE: PRT
/ ORGANISM: Zea mays

```

FEATURE:  
OTHER INFORMATION: Clone ID: UC-ZMFLB73265D07\_FLI.pcp  
US-10-425-114-68280

Query Match 26.3%; Score 448; DB 12; Length 482;  
Best Local Similarity 36.4%; Pred. No. 2,3e-33;  
Matches 123; Conservative 44; Mismatches 113; Indels 58; Gaps 10;

QY 3 AMRKALLAWYRENAAPLPWR-----GKDPYRVLVSEVLLQOTRVEQALPYRRFLERP 57  
DB ALAAQQLRMVDARHRLDPMRCVSGSSEERAYAVWVSEVLLQOTRVPVVAAYERMAWMP 129  
QY 58 TLKALAAASLEBVLRWQAGYRRARHLRLARVSE---LPPSFELGGLGLGYTA 114  
DB TVRSLAAATOEVEVEMWAGYRRARFLLEGAKQIIEKGLPCTALALAEVVGIGDYTA 189  
QY 115 AAVASTAFGERVAADVGNVRVLSRLFA-RESPKEKELFALAQGLPEGV---PGVWNO 170  
DB GALASTAFNEVVPVVDGNVIRVLSRLTYTIDNPKESSTVKRFMDLVQGMVDPLRPGDFNQ 249  
QY 171 ALMELGATVCLPKRPGCAGPLGAF-----RGEKAPGRYPAPR---KRAKEERLVAL 221  
DB 250 AMMELGATLSCSKTPGSCQPSVSHCOALALSRKSSVQVTDPRVVPKAKPRSDFAVVC 309  
QY 222 VLLGRKV-----HLERL-----ECRFQGLYGVLPFPE---ELPGREAFG- 260  
DB 310 VVQIQAQGLEBAADPKGNHDLFLIKRPEBGLAGLMEFPLVIVDQKTDLLNRKMDK 369  
QY 261 -----VRSRPLGEVRHALTFRRLRVEV 282  
DB 370 YLSKLSIDMVRKPDVILREDDVGHVHIFSHIRLTMV 407

RESULT 5

US-10-437-963-138271  
Sequence 138271, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 138271  
LENGTH: 474  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_39677C.1.pcp  
US-10-437-963-138271

Query Match 26.0%; Score 442; DB 16; Length 474;  
Best Local Similarity 35.0%; Pred. No. 8,3e-33;  
Matches 122; Conservative 47; Mismatches 110; Indels 70; Gaps 12;

QY 3 AMRKALLAWYRENAAPLPWRGKDP-----YRVLVSEVLLQOTRVEQALPY 48  
DB 56 AVRELLMWDNRDLPWRRAAEPPASSGSGRGEQAIVWVSEVLLQOTRVPVVVDY 115  
QY 49 YRRFLERPPTLKALAAASLEBVLRWQAGYRRARHLRLARV---EELPPSFALRG 105  
DB 116 YSRWMAAPTYDSLAAATOEVEVEMWAGLGYRRARFLLEGAKQIVKGEFPCTASTLR 175  
QY 106 LPLGLPYTAAVAASIAFERVAADVGNVRVLSRLFA-RESPKEKELFALAQGLPEGV 164

DB 176 VRGIGDYTGALASTAFNENVPVVDGNVRVLSRFYALPDNPKESSTVKRFMQLTDELVD 235  
QY 165 ---PGVWNOALMELGATVCLPKRPGCAGPLGAFGRKAPGR-----YR--APRR 211  
DB 236 PSRPGDFNOAMMELGATLSCSKTPGSCQPSVSHCOALALSSQMSVKVTDPRVVPKAK 295  
QY 212 RAKEERLVALVLLGRGVHLERLEGRFQGLYGVLPFPEELPGREAFGRSRPLGEVRH 271  
DB 296 PMSDFAVCVQVIS-----QGRFEGI-----AEAGKNDLFLIKRP----- 332  
QY 272 ALTHRLRIVEVRGALWE-----GEGE-DPMKRPLPKMEKYLKALPL 313  
DB 333 -----EGLAGLMEFPLVINEGKTDTLNR--RKENDKYLKQLLSI 372

RESULT 6

US-10-629-951-36  
Sequence 36, Application US/10629951  
Publication No. US20040018550A1  
GENERAL INFORMATION:  
APPLICANT: Bellacosa, Alfonso  
TITLE OF INVENTION: Methods for Detection of Transition  
TITLE OF INVENTION: Single-Nucleotide Polymorphisms  
FILE REFERENCE: FCCC 96-21  
CURRENT APPLICATION NUMBER: US/10/629,951  
CURRENT FILING DATE: 2003-07-29  
PRIOR APPLICATION NUMBER: US/09/629,222A  
PRIOR FILING DATE: 2000-07-31  
PRIOR APPLICATION NUMBER: 09/463,891  
PRIOR FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: PCT/US98/15828  
PRIOR FILING DATE: 1998-07-28  
PRIOR APPLICATION NUMBER: 60/053,936  
PRIOR FILING DATE: 1997-07-28  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 36  
LENGTH: 188  
TYPE: PRT  
ORGANISM: E. coli  
US-10-629-951-36

Query Match 22.9%; Score 389.5; DB 15; Length 188;  
Best Local Similarity 44.8%; Pred. No. 2,2e-28;  
Matches 81; Conservative 37; Mismatches 50; Indels 13; Gaps 5;

QY 27 PYRVLVSEVLLQOTRVEQALPYRRFLERPPTLKALAAASLEBVLRWQAGYRRARHL 86  
DB 2 PYKWLSEVLLQOTQVATVIFPERFMARPFYTDLANAPLDEVLHMTGLGYARARNT 61  
QY 87 HRLARVSEEL-----PPSFALRGPLGGLGYTAANAASIAFERVAADVGNVRVLSRLF 141  
DB 62 HKAAQVATLHGKFPETFEVVALPGVGRSTAGALISLKGKFPILDNVVKVRLARCY 121  
QY 142 ARRS-PKEK-----LFLAAGLPL- GVDGCVWNOALMELGATVCLPKRPGCAGPLGAF 195  
DB 122 AVSGMPKKEVENKLSLSQVTPAVGVER--FNQAMMDIGAMI CTRSKRKCSICPLONG 179  
QY 196 C 196  
DB 180 C 180

RESULT 7

US-10-156-761-12241  
Sequence 12241, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI

```

/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ CURRENT FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 12241
/ LENGTH: 313
/ TYPE: PRT
/ ORGANISM: Streptomyces avermitilis
US-10-156-761-12241

```

```

Query Match
Best Local Similarity 22.2%; Score 377; DB 14; Length 313;
Matches 97; Conservative 32; Mismatches 88; Indels 28; Gaps 7;

```

```

QY 2 EAMRKALLAWYRENARPLPM-RGEKDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLK 60
DB 25 EALHAQVIMFDEHNDLPMRRPDGPMGVWVSEEMLOOTPVNRVLPYEQWLMARMPRA 84
QY 61 ALAASLEEVLRVWQAGYRRRAEHLRLARSVEE-LPPSPFAELRGULPGIGPTAA 115
DB 85 DLAKERPGEAVRWAGGLGYPRRLRLHGAVALTERHNHGDVPTHEAQLALPGIGETAA 144
QY 116 AVASIAFGEVVAVDGNVRVLSRLP-----ARSPKKEKLFALAQGLPBGVDPGV 167
DB 145 AVASPAVYCGRAHVLDTNRVRFARAVTVGVPPNATTAERK-----LAPALLPEDESTAS 200
QY 168 -WQALMELGATVCLPKRPGCAGLGAFCRGKEA--PGRYPAPRK-----RRAKEER 217
DB 201 RMAAASMEIGALVCTAKNETCHRCPLNGQCAMRLAGLPBHDGPRKQGTAGTDRQVGRG 260
QY 218 LVALV 222
DB 261 LLAIV 265

```

## RESULT 8

```

US-09-738-626-6433
/ Sequence 6433, Application US/09738626
/ Publication No. US20020197605A1
/ GENERAL INFORMATION:
/ APPLICANT: NAKAGAWA, SATOSHI
/ APPLICANT: MIZOGUCHI, HIROSHI
/ APPLICANT: ANDO, SEIKO
/ APPLICANT: HAYASHI, MIKIRO
/ APPLICANT: OCHIAI, KEIKO
/ APPLICANT: YOKOI, HARUHIKO
/ APPLICANT: TATEISHI, NAKO
/ APPLICANT: SENOH, AKIHIRO
/ APPLICANT: IKEDA, MASATO
/ APPLICANT: OZAKI, AKIO
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-125
/ CURRENT APPLICATION NUMBER: US/09/738,626
/ CURRENT FILING DATE: 2000-12-18
/ PRIOR APPLICATION NUMBER: JP 99/377484
/ PRIOR FILING DATE: 1999-12-16
/ PRIOR APPLICATION NUMBER: JP 00/159162
/ PRIOR FILING DATE: 2000-04-07
/ PRIOR APPLICATION NUMBER: JP 00/280988
/ PRIOR FILING DATE: 2000-08-03
/ NUMBER OF SEQ ID NOS: 7059
/ SOFTWARE: PatentIn ver. 3.0
/ SEQ ID NO 6433
/ LENGTH: 293
/ TYPE: PRT

```

```

/ ORGANISM: Corynebacterium glutamicum
US-09-738-626-6433

```

```

Query Match
Best Local Similarity 21.5%; Score 365; DB 9; Length 293;
Matches 99; Conservative 47; Mismatches 120; Indels 20; Gaps 7;

```

```

QY 3 AMRKALLAWYRENARPLPMRGEKDP-----YRVLVSEVLLQOTRVEQALPYRRFLERFPT 58
DB 5 AFQTAALVWFRANARDLAAR---DPNTSAMGILLSVWSQOTPVARVPIRREWMERKPT 61
QY 59 LKALAAASLEEVLRVWQAGYRRRAEHLRLA-----RSVEELPPSPFAELRGULPGPYT 113
DB 62 PEDFANASTDDELIRSKGKGYPRRLRLKECAEYIVKHAQGEVDPTEALALALPGIGDYT 121
QY 114 AAASIAFGEVVAVDGNVRVLSRLPARE---SPKKEKLFALAQGLPBGVDPGVWQ 170
DB 122 ARAVAAFHGQGVVPVDTNVRVYQRAVAGRYLAGPAKQGLIDVSLLPNTAP--EFS 180
QY 171 ALMELGATVCLPKRPGCAGLGAFCRGKEAPGRYPAPRRRAKEERLVALVLGR--KG 228
DB 181 ALMELGALICTRTSPKCDTCPLDQCQWQKGCPSPESEELAAKRVQKFGVGTDRQVNG 240
QY 229 VHLERLEGRFO--GLYGVLPFPPEELPGRREAPFGRSRPLGEVRHA 272
DB 241 LIMDVLRNATAPVPLSAIDVWVPDDAQRSLRLSLIEDGLAEQNEA 286

```

## RESULT 9

```

US-10-437-963-103808
/ Sequence 103808, Application US/10437963
/ Publication No. US20040123343A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Wu, Wei
/ APPLICANT: Boukharov, Andrey A.
/ APPLICANT: Barbazuk, Brad
/ APPLICANT: Li, Ping
/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ CURRENT FILING DATE: 2003-05-14
/ NUMBER OF SEQ ID NOS: 204966
/ SEQ ID NO 103808
/ LENGTH: 1719
/ TYPE: PRT
/ ORGANISM: Oryza sativa
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(1719)
/ OTHER INFORMATION: unsure at all Xaa locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_101202C.1.pep
US-10-437-963-103808

```

```

Query Match
Best Local Similarity 19.8%; Score 337; DB 16; Length 1719;
Matches 105; Conservative 29; Mismatches 88; Indels 88; Gaps 11;

```

```

QY 3 AMRKALLAWYRENARPLPMR-----GEKDPYRVLVSEVLLQOTRVEQALP 47
DB 54 AVRAEELRLMYANRRDLPMBRAABPDAGSGSGRGEKAYVWVWSEVWLQHTRPVVVD 113
QY 48 YRRPFLERFPTLKAALAAASLEEVLRVWQAGYRRRAEHLRLARSV---EELPPSPFAELR 104
DB 114 YSRWMAWPTVDSIAAATGEEVNMWAGLGYRRRRLLEGAQKQIVEGGEFRTASALR 173
QY 105 GLPGLGPTAAVAASIAFGEVVAVDGNVRVLSRLPFA--RESPKKEKLFALAQGLPBGV 163

```



Db 174 EVRGIDYTAGAIAIAFNEVVPVVGDNV-RVLSRLYALPDNKE----- 217  
 QY 164 DPGVMQALMELGATVCLPRPRGACPLGAFCKRGAPGRYPAPRRR-----A 213  
 Db 218 -----SSTV---KR-----FCRRQDAELAAAEERRRATARTGMAAA 252  
 QY 214 KEERLVALVLGRKG-----VHLERLEGRQGLGVPLFPPELPGREAA 258  
 Db 253 RAARLAAAEIAAARAEEAEADAARAAAEVEVETLSSINGSIAGDITADREL--EELA 310  
 QY 259 FGVRSPRLGE 268  
 Db 311 RG-RARERRAE 319

## RESULT 10

US-10-424-599-248081  
 ; Sequence 248081, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 248081  
 ; LENGTH: 184  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_66047C.1.pcp  
 ; US-10-424-599-248081

Query Match 15.5%; Score 263.5; DB 12; Length 184;  
 Best Local Similarity 45.5%; Pred. No. 1.6e-16;  
 Matches 56; Conservative 19; Mismatches 35; Indels 13; Gaps 2;

QY 5 RKALLANYRENAPLPRG-----EKDPYVLYSEVLQOTRYEOLPYRRRLER 55  
 Db 62 RVALLDWYDINRRDLPRRTTFKQDEEVEERAYGVWVSEVWLQOTRVQTVIAYYNRWQK 121  
 QY 56 PPTLKALAAASLEEVLVWOGAGYRRAEHLRLARV---EELPPSPALRGLPGIGP 111  
 Db 122 WPTIHLAQAASLEEVNEMWAGLYRRARFLLEGAKKI VAEGGQIPKVASMLRNIPGIGE 181  
 QY 112 YTA 114  
 Db 182 YTS 184

## RESULT 11

US-10-629-951-34  
 ; Sequence 34, Application US/10629951  
 ; Publication No. US20040018550A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bellacosa, Alfonso  
 ; TITLE OF INVENTION: Methods for Detection of Transition  
 ; TITLE OF INVENTION: Single-Nucleotide Polymorphisms  
 ; FILE REFERENCE: FCCC 96-21  
 ; CURRENT APPLICATION NUMBER: US/10/629,951  
 ; CURRENT FILING DATE: 2003-07-29  
 ; PRIOR APPLICATION NUMBER: US/09/629,222A  
 ; PRIOR FILING DATE: 2000-07-31  
 ; PRIOR APPLICATION NUMBER: 09/463,891  
 ; PRIOR FILING DATE: 2000-01-28  
 ; PRIOR APPLICATION NUMBER: PCT/US98/15828  
 ; PRIOR FILING DATE: 1998-07-28  
 ; PRIOR APPLICATION NUMBER: 60/053,936

; PRIOR FILING DATE: 1997-07-28  
 ; NUMBER OF SEQ ID NOS: 73  
 ; SOFTWARE: FASTSEQ for Windows Version 3.0  
 ; SEQ ID NO 34  
 ; LENGTH: 188  
 ; TYPE: PRT  
 ; ORGANISM: M. thermophilicum  
 ; US-10-629-951-34

Query Match 14.1%; Score 240.5; DB 15; Length 188;  
 Best Local Similarity 32.0%; Pred. No. 2.4e-14;  
 Matches 58; Conservative 35; Mismatches 77; Indels 11; Gaps 4;

QY 26 DPNRVVSEVLQOTRYEOLPYRRRLERFPTLKALAAASLEEVLVWOGAGY-VRRAE 84  
 Db 1 DPNVILITELIRRTTAGHVKKIYDFEVKTCFEDILTPKSEIKDKIETGLSNQRAE 60  
 QY 85 HLHRLARV---EELPPSPALRGLPGIGPYTAAVAASIAFGEVAAVVDGNVRVLSR 139  
 Db 61 QLKELARVINDYGVPRNRKAIIDLPGVGKKTCAAVWGLAFGKKAADVAVFVAVINR 120  
 QY 140 LFARE---SPKKEFLPALAQGLLPGVDPGVWQALMELGATVCLPKRPRGACPLGAF 195  
 Db 121 YFGSYENINYNHKAALMELAEITLVPQKCRD-FNLGLMDFSAITCAPRRKCKEKGMSKL 179  
 QY 196 C 196  
 Db 180 C 180

## RESULT 12

US-10-335-977-7283  
 ; Sequence 7283, Application US/10335977  
 ; Publication No. US20040052799A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DOUGLAS SMITH et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
 ; RELATING TO HELICOBACTER PYLORI FOR  
 ; DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHIVE & COCKFIELD  
 STREET: 28 State Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109-1875

COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible  
 OPERATING SYSTEM: Windows NT 4.0  
 SOFTWARE: UNIX

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002  
 FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:  
 NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400  
 TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 7283:  
 SEQUENCE CHARACTERISTICS:

LENGTH: 230 amino acids  
 TYPE: amino acid

TOPOLOGY: linear  
 MOLECULE TYPE: protein

HYPOTHETICAL: YES  
 ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori  
 NAME/KEY: misc feature  
 LOCATION: (B) LOCATION 1...230  
 SEQUENCE DESCRIPTION: SEQ ID NO: 7283:  
 US-10-335-977-7283

Query Match 13.8%; Score 234; DB 12; Length 230;  
 Best Local Similarity 30.6%; Pred. No. 1.3e-13;  
 Matches 66; Conservative 32; Mismatches 54; Indels 64; Gaps 9;

QY 95 ELPPSPFAELGRLPGCPYTAATAVAASIAFGERYAAVDGVRVLSRLFARSPKKEKLFAL 154  
 DB 7 QLPNDYQSLLKPLGIGAYTANAALCFGRKSAACVDANVKVLRPF----- 53  
 QY 155 AAGLLPBGVDPCV-----WQALMELGATVCLPKRPGACPLGA 194  
 DB 54 -----GLDPNTHAKDQIKANDPLNLSFNHQAALIDGALICSPK-PKCAICPFNP 105  
 QY 195 FCRGKEAPGRYPAPRRK-AREERLVALVLLGRGVHLERLEGRFOGLY-GVPLFP----- 248  
 DB 106 YCLGNHLERHTLKKKQEIIOERYLGVVIONNQ-IALEKIE---QKLYIGMHHPNKE 161  
 QY 249 --PEELPGRAAFGRSRPLGEVHALTHRLAVEV 282  
 DB 162 NLEFKLPF-----LGTIKHSHTKFLMLNL 186

# RESULT 13

US-10-335-977-7282  
 ; Sequence 7282, Application US/10335977  
 ; Publication No. US20040052799A1  
 ; GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
 RELATING TO HELICOBACTER PYLORI FOR  
 DIAGNOSTICS AND THERAPEUTICS  
 NUMBER OF SEQUENCES: 10031  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHIVE & COCKFIELD  
 STREET: 28 State Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109-1875

COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: IBM PC Compatible  
 OPERATING SYSTEM: Windows NT 4.0  
 SOFTWARE: UNIX

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/335,977  
 FILING DATE: 30-Dec-2002  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/993,002  
 FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:  
 NAME: Mandragouras, Amy B.  
 REGISTRATION NUMBER: 36,207  
 REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)227-7400  
 TELEFAX: (617)742-4214  
 INFORMATION FOR SEQ ID NO: 7282:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 187 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Helicobacter pylori

FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (B) LOCATION 1...230  
 SEQUENCE DESCRIPTION: SEQ ID NO: 7283:  
 US-10-335-977-7283

NAME/KEY: misc feature  
 LOCATION: (B) LOCATION 1...187  
 SEQUENCE DESCRIPTION: SEQ ID NO: 7282:  
 US-10-335-977-7282

Query Match 13.4%; Score 228.5; DB 12; Length 187;  
 Best Local Similarity 33.5%; Pred. No. 3.3e-13;  
 Matches 59; Conservative 25; Mismatches 45; Indels 47; Gaps 7;

QY 95 ELPPSPFAELGRLPGCPYTAATAVAASIAFGERYAAVDGVRVLSRLFARSPKKEKLFAL 154  
 DB 16 QLPNDYQSLLKPLGIGAYTANAALCFGRKSAACVDANVKVLRPF----- 62  
 QY 155 AAGLLPBGVDPCV-----WQALMELGATVCLPKRPGACPLGA 194  
 DB 63 -----GLDPNTHAKDQIKANDPLNLSFNHQAALIDGALICSPK-PKCAICPFNP 114  
 QY 195 FCRGKEAPGRYPAPRRK-AREERLVALVLLGRGVHLERLEGRFOGLY-GVPLFP 248  
 DB 115 YCLGNHLERHTLKKKQEIIOERYLGVVIONNQ-IALEKIE---QKLYIGMHHP 166

# RESULT 14

US-10-335-977-7279  
 ; Sequence 7279, Application US/10335977  
 ; Publication No. US20040052799A1  
 ; GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
 RELATING TO HELICOBACTER PYLORI FOR  
 DIAGNOSTICS AND THERAPEUTICS  
 NUMBER OF SEQUENCES: 10031  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHIVE & COCKFIELD  
 STREET: 28 State Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109-1875

COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: IBM PC Compatible  
 OPERATING SYSTEM: Windows NT 4.0  
 SOFTWARE: UNIX

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/335,977  
 FILING DATE: 30-Dec-2002  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/993,002  
 FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:  
 NAME: Mandragouras, Amy B.  
 REGISTRATION NUMBER: 36,207  
 REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)227-7400  
 TELEFAX: (617)742-4214  
 INFORMATION FOR SEQ ID NO: 7279:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 98 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Helicobacter pylori

FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (B) LOCATION 1...98  
 SEQUENCE DESCRIPTION: SEQ ID NO: 7279:  
 US-10-335-977-7279

Query Match 11.8%; Score 200.5; DB 12; Length 98;



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 6, 2004, 20:11:37 ; Search time 19 Seconds  
(without alignments)  
883.076 Million cell updates/sec

Title: US-09-938-901a-2

Sequence: 1 MEAWRKALLAWYRENRPLP.....VLRKALPLAHGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	505	29.7	403	4 US-09-107-532A-7308	Sequence 7308, Ap
2	485	28.5	395	4 US-09-134-000C-5115	Sequence 5115, Ap
3	481.5	28.3	535	3 US-08-813-574-2	Sequence 2, Appl1
4	477.5	28.1	470	4 US-09-252-991A-2310	Sequence 2310, A
5	477	28.0	375	4 US-09-198-452A-421	Sequence 421, App
6	471.5	27.7	350	3 US-08-813-574-9	Sequence 9, Appl1
7	471.5	27.7	350	4 US-09-651-656-37	Sequence 37, Appl1
8	471.5	27.7	350	4 US-09-650-855-37	Sequence 37, Appl1
9	457.5	26.9	516	4 US-09-489-039A-11959	Sequence 11959, A
10	450	26.5	347	4 US-09-543-681A-5315	Sequence 5315, Ap
11	443	26.0	360	4 US-09-134-001C-4684	Sequence 4684, Ap
12	442	26.0	346	4 US-09-328-352-4982	Sequence 4982, Ap
13	389.5	22.9	188	4 US-09-629-222A-36	Sequence 36, Appl1
14	355	20.9	414	4 US-09-540-236-2578	Sequence 2578, Ap
15	274.5	16.1	221	1 US-08-663-023-17	Sequence 17, Appl1
16	274.5	16.1	221	1 US-09-402-959A-10	Sequence 10, Appl1
17	274.5	16.1	225	4 US-10-037-927B-14	Sequence 14, Appl1
18	240.5	14.1	188	4 US-09-629-222A-34	Sequence 34, Appl1
19	191	11.2	185	4 US-09-629-222A-35	Sequence 35, Appl1
20	188	11.1	227	4 US-09-107-532A-3843	Sequence 3843, Ap
21	176	10.3	207	2 US-08-808-550-36	Sequence 36, Appl1
22	176	10.3	259	2 US-08-808-550-39	Sequence 39, Appl1
23	172	10.1	209	2 US-08-808-550-31	Sequence 31, Appl1
24	171	10.1	204	2 US-08-808-550-32	Sequence 32, Appl1
25	164.5	9.7	211	2 US-08-808-550-37	Sequence 37, Appl1
26	164.5	9.7	297	2 US-08-808-550-42	Sequence 42, Appl1
27	164.5	9.7	304	2 US-08-808-550-2	Sequence 2, Appl1

28	160	9.4	224	4 US-09-134-001C-2225	Sequence 3225, Ap
29	158	9.3	185	4 US-09-134-000C-3440	Sequence 3440, Ap
30	154	9.1	213	2 US-08-808-550-35	Sequence 35, Appl1
31	145	8.5	207	2 US-08-808-550-29	Sequence 29, Appl1
32	145	8.5	211	2 US-08-808-550-38	Sequence 38, Appl1
33	145	8.5	211	2 US-09-651-656-25	Sequence 25, Appl1
34	145	8.5	211	4 US-09-650-855-25	Sequence 25, Appl1
35	142	8.3	216	4 US-09-489-039A-9560	Sequence 9560, Ap
36	137.5	8.1	184	4 US-09-629-222A-33	Sequence 33, Appl1
37	134	7.9	207	2 US-08-808-550-30	Sequence 30, Appl1
38	128	7.5	628	4 US-09-252-991A-30904	Sequence 30904, A
39	127.5	7.5	212	4 US-09-252-991A-17572	Sequence 17572, A
40	121	7.1	216	4 US-09-543-681A-4806	Sequence 4806, Ap
41	121	7.1	4472	2 US-08-804-227C-2	Sequence 2, Appl1
42	118.5	7.0	776	4 US-09-252-991A-28446	Sequence 28446, A
43	117	6.9	230	4 US-09-328-352-5770	Sequence 5770, Ap
44	115.5	6.8	494	4 US-09-252-991A-17295	Sequence 17295, A
45	115	6.8	542	4 US-09-252-991A-21753	Sequence 21753, A

# ALIGNMENTS

RESULT 1  
US-09-107-532A-7308  
Sequence 7308, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: US/09/107,532A  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-8277  
TELEFAX: (781)893-5007  
INFORMATION FOR SEQ ID NO: 7308:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
MOLECULE TYPE: linear  
TOPOLGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) Location 1...403  
SEQUENCE DESCRIPTION: SEQ ID NO: 7308:  
US-09-107-532A-7308



Db 260 RPDGDFQAAWELGATVCTPQRPICSCQPVESLCRAQRYEQEOLLASGSLSGSDVEECA 319  
 QY 202 P-----GRYPAPRKRAK---ERLVALVL-----LGRKVHLER- 233  
 Db 320 PNTGQCHLCIPSEPMQDTLGVNFRPKRSRKPRESSATCTLEQDGLGAILLVQRP 379  
 QY 234 LSGRFQGLYGP--LPPPEELPGRBAF-----GVSRPPIGEVRAHLTHRLRY 280  
 Db 380 NSGLLGLWEPFSPVTWEPSEQLQKALLQELQWAGPLPATHLRHLEVVHTFSHAKLTY 439  
 QY 281 EYRGALMEGE 250  
 Db 440 QVYGLALEGQ 449

RESULT 4  
 US-09-252-991A-23310  
 ; Sequence 23310, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 23310  
 ; LENGTH: 470  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-23310

Query Match 28.1%; Score 477.5; DB 4; Length 470;  
 Best Local Similarity 36.2%; Pred. No. 1.3e-39;  
 Matches 126; Conservative 58; Mismatches 119; Indels 45; Gaps 14;

QY 2 EAMRKALLAWY-RENAAPLPWGEKDPYRVLVSEVLLQOTRVEQALPYRRFELERFPTLK 60  
 Db 119 EGFNGAVLDWYDRGRXDLPWQOQITPRVWVSEIMLQOTVSTVLGFDPMALPDVE 178  
 QY 61 ALAASLEEVLRVWQAGYVRAHLEHLA-----RSVEELPSPFAELRGLPGGYTAA 115  
 Db 179 ALAASAEDEVLAHMTGLGYISRAKMLHKTQIVVERHAGEPFRVDEQLAELPGIGRSTAG 238  
 QY 116 AVASIAFGERVAADVGNVRVLSRLFARE---SPK-EKELFALAQGLPBGVDPGYWNO 170  
 Db 239 AIAHLSNGLRAPILIDGNVRKVLARYLAQDYGPGSEKVARALWMAERTPH-ARVNHYTQ 297  
 QY 171 ALMELGATVCLPKPRCGACPLGAFPCR---GKEAPGRYPAPRKRA-KERLVALVLIG 225  
 Db 298 AMMDLGATLCTRSRPSCLCPVSGCRALHGREAD--YPOPKRKALPQKRTIMPILAN 355  
 QY 226 RKG---VHLERLEGRFOGLYGVPLFPPEELPGRBAFVGRSRPGEV-----HALTHR 276  
 Db 356 RDGAILLYRRPSSGLWGLMSLPBL--DDLGLLEPLAARHSLALGERRELSGLTHHTSHF 413  
 QY 277 RLRYVE-----VRGA-----LWEGEGEDPWKRPLPKMEKYLKKA 310  
 Db 414 QLAIEPMIVAVEGA PRAVABGDWMLVYNLATPP-RLGGLAAYPKLKLKRA 460

RESULT 5  
 US-09-198-452A-421  
 ; Sequence 421, Application US/09198452A  
 ; Patent No. 6559294  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Griffiths, R.  
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve  
 ; TITLE OF INVENTION: and treatment of infection  
 ; FILE REFERENCE: 9710-003-999  
 ; CURRENT APPLICATION NUMBER: US/09/198,452A  
 ; CURRENT FILING DATE: 1998-11-24  
 ; NUMBER OF SEQ ID NOS: 6849  
 ; SEQ ID NO 421  
 ; LENGTH: 375  
 ; TYPE: PRT  
 ; ORGANISM: Chlamydia pneumoniae  
 ; US-09-198-452A-421

Query Match 28.0%; Score 477; DB 4; Length 375;  
 Best Local Similarity 38.6%; Pred. No. 1.1e-39;  
 Matches 117; Conservative 44; Mismatches 112; Indels 30; Gaps 7;

QY 6 KALLAWYRENAAPLPWGEKDPYRVLVSEVLLQOTRVEQALPYRRFELERFPTLKALAA 65  
 Db 22 BALKKWFENKXKSLPFRDNPPTPYSVWVSEVMLQOTRAEVVIDYFNQWMEFPPIESLAA 81  
 QY 66 SLEEVLRVWQAGYVRAHLEHLASVEE-----LPPFAELRGLPGGYTAAVASI 120  
 Db 82 KEEDVIXKMEGIGYSRARHLLEGARWVEEFHGKIPDDAISLAQIRGVGPTVHALIAP 141  
 QY 121 AGERVAADVGNVRVLSRLFARESPEKE-----LFALAQGLPBGVDPGYWNOALMBL 175  
 Db 142 AFRRAAADVGNVLRVLSRIFLIETISDLESTRVWSRIQALLPH-KSEVIAEHLIEL 200  
 QY 176 GATVCLPKPRCGACPLGAFPCRGEAPGRYPAPRKRAK---ERLVALVL-GRKGVH 230  
 Db 201 GACIC-KKYPQCHRCVVRQACGAMRENKQFVLPVRARKKVIPLHRLVALVLDGSLVVE 259  
 QY 231 LERLEGRFOGLYGVPLFPPEELPGRBAFVGRSR-----PLGEVHALTHRR 277  
 Db 260 KRRPKEMAGLYEFPYIEVEPEBGLDIEGFTKMLSLSPLEFLGNLKEQHAAFTNHK 319  
 QY 278 LRV 280  
 Db 320 VHL 322

RESULT 6  
 US-08-813-574-9  
 ; Sequence 9, Application US/08813574  
 ; Patent No. 6013473  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wei, Ying-Fei  
 ; TITLE OF INVENTION: Human Mucy  
 ; NUMBER OF SEQUENCES: 18  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Smithkline Beecham Corporation  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19406  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: Fastseq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/813,574  
 ; FILING DATE:  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/013,132  
 ; FILING DATE: 11-MAR-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Han, William T  
 ; REGISTRATION NUMBER: 34,344  
 ; REFERENCE/DOCKET NUMBER: ATG50002  
 ; TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5219  
 TELEFAX: 610-270-4026  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 350 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-813-574-9

Query Match 27.7%; Score 471.5; DB 3; Length 350;  
 Best Local Similarity 35.1%; Pred. No. 3.5e-39;

Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAATRENNAR-PLPWREGKDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
 DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVWLQOTVAVIIPYFERFMARFPTVLDLANAP 69  
 QY 67 LEEVLRVWQAGYRRAEHLRLASVEEL-----PSPFAELRGIPGLGYTAATAVASIA 121  
 DB 70 LDEVLHLMTGLGYTAARNLHKAQOVATLHGCKPEFTEEEVVALPGVGRSTAGAILSL 129  
 QY 122 FGERVAADVGNVRRVLSLFARES--PKKE-----LFALAOGILPE-GVDPGVNQAIME 175  
 DB 130 LGHGFPIIDGNVRYLARCVAASGWPCKKEVENKMLMSIEQVTPAVGER--FNQAMMDL 187  
 QY 176 GATVCLPKRPGCAGPL--GAFCRGKEAPGRYPARKRAKEERLVALVLLGRGVHL-- 231  
 DB 188 GAMICTRSKPKCSLCPLONGCIAAANNSWALYPGKKPKQTLPERTGYFLLQHEDEVLLA 247  
 QY 232 -ERLEGRFGILYGVLPFPEE-----LPGREAFGVSRPLGEVRHALTHRLRV----- 280  
 DB 248 QRPSPGLMGILYCPQFADBSLRQWLAQRQIADNLTQ-LTAFRHTFSHFHLDIVPMWL 306  
 QY 281 -----EVRGALMBEGEDP---WKRPLPKLMEKVLRAKALPL 313  
 DB 307 PVSSFTGCMDEGNALMYNLAQPPSVGLAAPERLLQO-LRTGAPV 350

RESULT 7

US-09-651-656-37  
 Sequence 37, Application US/09651656  
 Patent No. 6340566  
 GENERAL INFORMATION:  
 APPLICANT: MCUTHERN-MALONEY, SANDRA  
 TITLE OF INVENTION: LAMRENCE LIVERMORE NATIONAL LABORATORY  
 TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE  
 POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,  
 FILE REFERENCE: IL-10689  
 CURRENT APPLICATION NUMBER: US/09/651,656  
 PRIOR FILING DATE: 2000-08-29  
 PRIOR APPLICATION NUMBER: 60/192,764  
 NUMBER OF SEQ ID NOS: 106  
 SOFTWARE: Patent In Ver. 2.1  
 SEQ ID NO 37  
 LENGTH: 350  
 TYPE: PRN  
 ORGANISM: Escherichia coli  
 US-09-651-656-37

Query Match 27.7%; Score 471.5; DB 4; Length 350;  
 Best Local Similarity 35.1%; Pred. No. 3.5e-39;

Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAATRENNAR-PLPWREGKDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
 DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVWLQOTVAVIIPYFERFMARFPTVLDLANAP 69  
 QY 67 LEEVLRVWQAGYRRAEHLRLASVEEL-----PSPFAELRGIPGLGYTAATAVASIA 121

DB 70 LDEVLHLMTGLGYTAARNLHKAQOVATLHGCKPEFTEEEVVALPGVGRSTAGAILSL 129  
 QY 122 FGERVAADVGNVRRVLSLFARES--PKKE-----LFALAOGILPE-GVDPGVNQAIME 175  
 DB 130 LGHGFPIIDGNVRYLARCVAASGWPCKKEVENKMLMSIEQVTPAVGER--FNQAMMDL 187  
 QY 176 GATVCLPKRPGCAGPL--GAFCRGKEAPGRYPARKRAKEERLVALVLLGRGVHL-- 231  
 DB 188 GAMICTRSKPKCSLCPLONGCIAAANNSWALYPGKKPKQTLPERTGYFLLQHEDEVLLA 247  
 QY 232 -ERLEGRFGILYGVLPFPEE-----LPGREAFGVSRPLGEVRHALTHRLRV----- 280  
 DB 248 QRPSPGLMGILYCPQFADBSLRQWLAQRQIADNLTQ-LTAFRHTFSHFHLDIVPMWL 306  
 QY 281 -----EVRGALMBEGEDP---WKRPLPKLMEKVLRAKALPL 313  
 DB 307 PVSSFTGCMDEGNALMYNLAQPPSVGLAAPERLLQO-LRTGAPV 350

RESULT 8

US-09-650-855-37  
 Sequence 37, Application US/09650855  
 Patent No. 6363555  
 GENERAL INFORMATION:  
 APPLICANT: MCUTHERN-MALONEY, SANDRA  
 TITLE OF INVENTION: LAMRENCE LIVERMORE NATIONAL LABORATORY  
 TITLE OF INVENTION: CHEMICAL PROTEINS FOR DETECTION AND QUANTITATION OF DNA  
 MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA  
 FILE REFERENCE: IL-10284  
 CURRENT APPLICATION NUMBER: US/09/650,855  
 PRIOR FILING DATE: 2000-08-29  
 PRIOR APPLICATION NUMBER: 60/192,764  
 NUMBER OF SEQ ID NOS: 106  
 SOFTWARE: Patent In Ver. 2.1  
 SEQ ID NO 37  
 LENGTH: 350  
 TYPE: PRN  
 ORGANISM: Escherichia coli  
 US-09-650-855-37

Query Match 27.7%; Score 471.5; DB 4; Length 350;  
 Best Local Similarity 35.1%; Pred. No. 3.5e-39;

Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAATRENNAR-PLPWREGKDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
 DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVWLQOTVAVIIPYFERFMARFPTVLDLANAP 69  
 QY 67 LEEVLRVWQAGYRRAEHLRLASVEEL-----PSPFAELRGIPGLGYTAATAVASIA 121  
 DB 70 LDEVLHLMTGLGYTAARNLHKAQOVATLHGCKPEFTEEEVVALPGVGRSTAGAILSL 129  
 QY 122 FGERVAADVGNVRRVLSLFARES--PKKE-----LFALAOGILPE-GVDPGVNQAIME 175  
 DB 130 LGHGFPIIDGNVRYLARCVAASGWPCKKEVENKMLMSIEQVTPAVGER--FNQAMMDL 187  
 QY 176 GATVCLPKRPGCAGPL--GAFCRGKEAPGRYPARKRAKEERLVALVLLGRGVHL-- 231  
 DB 188 GAMICTRSKPKCSLCPLONGCIAAANNSWALYPGKKPKQTLPERTGYFLLQHEDEVLLA 247  
 QY 232 -ERLEGRFGILYGVLPFPEE-----LPGREAFGVSRPLGEVRHALTHRLRV----- 280  
 DB 248 QRPSPGLMGILYCPQFADBSLRQWLAQRQIADNLTQ-LTAFRHTFSHFHLDIVPMWL 306  
 QY 281 -----EVRGALMBEGEDP---WKRPLPKLMEKVLRAKALPL 313  
 DB 307 PVSSFTGCMDEGNALMYNLAQPPSVGLAAPERLLQO-LRTGAPV 350

RESULT 9





```

; Sequence 4982, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAYMANNI FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4982
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4982

```

```

Query Match          26.0%; Score 442; DB 4; Length 346;
Best Local Similarity 33.3%; Pred. No. 3.2e-36;
Matches 113; Conservative 60; Mismatches 118; Indels 48; Gaps 11;

```

```

QY 7 ALAATRENAR-ELPWRGKDPYRLVSEVLLQOTRVEQALPYRRFLERPEPTLKALAA 65
DB 11 ALNMFDOHGRHDLPMQVADDPKVVWSEIMLQOTQVKTIVLYFRMERPEPTVEALGYA 70
QY 66 SIEEVLRVWQAGYVYARAHILRLARSVE--LPSPFAELRGLPGLPYTAAVASIAF 122
DB 71 TDEVAIPYAGLGTYARARNILHKAAGLVAAQOKFPEETLEEMIALPGIGSTAGALMSGL 130
QY 123 GERVAADVGNVRVLSRLFARES-----PKEKEFLAAGLLPESVDGVNQAALMELGA 177
DB 131 ROYGVIMDGNVKKVLAFFPAIEDLSKQHEREMKALAEELCTHRNND--TTQALMDLGA 189
QY 178 TWCLPRPRGACGACGAF---RKEKAPRGYRPRKRRAKEELVALVLGRKVALER 233
DB 190 TICTPKKPLCYLCPYQAHQAQVYQGLDELPFKPKTKTPYKADVLIIQCEDEWPMQQR 249
QY 234 -LEGFFQGLYGVLPPEPE-ELPGREAAFGVSRP-LGEVNHALLH----- 275
DB 250 QAHGIMGGLFCLPILENEHERLKLSQGFKQPOQOTQISHFTWLMNAHVFEVD 309
QY 276 --RLRVNVRGALMEGEGEDPWKRPLPKLMEKVLKALP 312
DB 310 QKEHLAIELEG-----QWLSP-----EOAIKGVF 334

```

```

RESULT 13
US-09-629-222A-36
; Sequence 36, Application US/09629222A
; Patent No. 6599700
; GENERAL INFORMATION:
; APPLICANT: Bellacosa, Alfonso
; TITLE OF INVENTION: Methods for Detection of Transition
; FILE REFERENCE: Single-Nucleotide Polymorphisms
; CURRENT APPLICATION NUMBER: US/09/629,222A
; CURRENT FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/463,891
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US98/15828
; PRIOR FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 60/053,936
; PRIOR FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FaestSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 188
; TYPE: PRT
; ORGANISM: E. coli
US-09-629-222A-36

```

```

Query Match          22.9%; Score 389.5; DB 4; Length 188;
Best Local Similarity 44.8%; Pred. No. 2.6e-31;
Matches 81; Conservative 37; Mismatches 50; Indels 13; Gaps 5;

```

```

QY 27 PYRLVSEVLLQOTRVEQALPYRRFLERPEPTLKALAAASLEEVLRVWQAGYVYARAH 86
DB 2 PYKWLSEVWLQOTQVATVPIPERFMARPEPTVTDLANPDLVEVHLMTGLGYARARNL 61
QY 87 HRLARSVEE-----PPSFAELRGLPGLPYTAAVASIAFGERVAADVGNVRVLSRLF 141
DB 62 HKAQVATVHLGKFPETPEBEVAALPGVGRSTAGAILSLISGKHPPIIDGVNKRVLARCY 121
QY 142 ARS-PKEKE---LFAAAGLLPE-GVDGVNQAALMELGATVCLPKRPGACGACPLGAF 195
DB 122 AVSGWPKKEVENKMLSEQVTPAVGYER--FNQAMMDLGMICTRSPKSCSLCPIDNG 179
QY 196 C 196
DB 180 C 180

```

```

RESULT 14
US-09-540-236-2578
; Sequence 2578, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2578
; LENGTH: 414
; TYPE: PRT
; ORGANISM: M. catarrhalis
US-09-540-236-2578

```

```

Query Match          20.9%; Score 355; DB 4; Length 414;
Best Local Similarity 28.5%; Pred. No. 2.3e-27;
Matches 109; Conservative 57; Mismatches 132; Indels 84; Gaps 13;

```

```

QY 3 AMRKALLAMYRENAR-PLPWRGKDP---YRLVSEVLLQOTRVEQALPYRRFLERPE 57
DB 33 SFARILLTWPELHGRHGLPMQYHNPASADIVANVWSEIMLQOTQVTVLKEFEPTLAFRA 92
QY 58 TLKALAAASLEEVLRVWQAGYVYARAHILRLARSVE-----LPSPFAELRGLPGLG 110
DB 93 TVGELAVADWQEVASFAGLGTYARARNILHKAAGLVAAQOVADFDTHGRFPETVEMQAVKVG 152
QY 111 PYTAAVASIAFGERVAADVGNVRVLSRLFA-----RESPKEKEFLAAGLLPEGVDP 165
DB 153 RSTAGAILVMGVYKFGVICDGNVKKVLAHRNAVCGDITKSAIDKRLMELATALTPEKYS- 211
QY 166 GVNQAALMELGATVCLPKRPRGACPLGAFCSGKEAPRGYRPRKRRAKEE---RLVALV 222
DB 212 GHYAQMMDLGAATICTTQPKCHLCPVTDDCAVALGVQSLPVKKAKAPRPHNSIALS 271
QY 223 LIG---RKGVHLEREGFQGLYGVLPPEPELP-----GREAA 258
DB 272 LTHGGLTWLHRQNGGGLWIDGIMSLPT--MLPLDNDOKLDNNTLSDAIFKAWQSDKV 328
QY 259 FGVRSRPLGEV-----RHALLH-----RLRVEVGA--- 285
DB 329 HDLHSQILEILPIPTQTLTAVALYRHTLVHMLYGMSTCIANNSQFNQIGTLTGIGIY 388
QY 286 LMEGEGEDPWKRPLPKLMEKVL 307
DB 389 LWT---DTPVNLPLPAMMKLL 407

```

```

RESULT 15
US-08-663-023-17
; Sequence 17, Application US/08663023
; Patent No. 5763178

```

## GENERAL INFORMATION:

APPLICANT: CHIRIKIAN, Jack G.  
 APPLICANT: COLLIER, G. Bruce  
 TITLE OF INVENTION: OSCILLATING SIGNAL AMPLIFIER FOR NUCLEIC  
 TITLE OF INVENTION: ACID DETECTION  
 NUMBER OF SEQUENCES: 17  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W., Suite 500  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/663,023  
 FILING DATE: 07-JUN-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/012,950  
 FILING DATE: 06-MAR-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/483,089  
 FILING DATE: 07-JUN-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 66669/110  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 17:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 221 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-663-023-17

Query Match 16.1%; Score 274.5; DB 1; Length 221;

Best Local Similarity 31.7%; Pred. No. 1.2e-19;

Matches 63; Conservative 40; Mismatches 85; Indels 11; Gaps 4;

```

QY      8 LLAWYRENARPLPVRGSKDPYIVSEVLLQOTRVEQALPYRRFLERFPTLKALAAASL 67
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     16 ILTFMNTDRDPRHRTDPPYVILITELRLRTTAGHVKKIYDKFVYKCFEDILKTPX 75
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      68 BEVLRWQAGCY-YRRAEHLRLARSV-----ELPPSFAELRGLPGLGPTAAVASIA 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      76 SEIAKDKEIGLSNQRAEQLELARVINDYGRVPRNRKAILDLPGVGYTCAAVMCLA 135
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     122 FGEERVAVDGVRVRVRLPARE---SPKEKELFALAQGLPBGVDPGVVNOALMELGA 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     136 FGKAAAVDANFVIVIRYFGSYENANVNHKALMELAEITVPGKCRD--FNLGLMDFSA 194
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      178 TVCLPKRPRCGACPLGAF 196
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     195 IICAPRRKCKEKGMSKLC 213
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  
```

Search completed: October 6, 2004, 20:19:07  
 Job time : 21 secs

**This Page Blank (uspto)**







RESULT 5  
US-09-925-301-484  
; Sequence 484, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 484  
; LENGTH: 1878  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1446)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-301-484

Query Match 10.7%; Score 104; DB 9; Length 1878;  
Best Local Similarity 54.0%; Pred. No. 2,1e-14;  
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

```
QY 69 GAGAGAGACCTTACCGCGCTGCTCGAGTCTCTTGAGAGACCGCGGTGGA 128
DB 509 GAGACGGGGGCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 568
QY 129 GCAAGCCCTCCCTATTACCGCGCTTCTTGAGACCGCTTCCACCTGAGGCGCTGCG 188
DB 569 CACTGTATCAACTACTATACCGGATGATGACAGAAATGCGCTTACACTGACAGACCTGCG 628
QY 189 CGCGGCTTCCCTGGAAGAGTCTTGTAGGTCTGCGAGGGGCGGGCTTACTACCGGGCGGC 248
DB 629 CAGTGTCTTCCCTGAGAGAGGTGATCACTTGGGCTGGGCTGCTGCTTACTATTCTGTGG 688
QY 249 GGAACACCTTCCACCGCGCTGCGCGGAGCGTGAAGAGCTTCCCGAGCTTCTC----- 300
DB 689 CCGGCGGCTGAGAGAGGAGCTCGAAGAGGTGTGAAGAGCTAGGGGGCCACATGCCACG 748
QY 301 -----GCCGAGCTTGGGGGCTTCTGTGTCTCGGGCTTACACCGCGGGCGGT 350
DB 749 TACAGACAGAGACCTTGAGAGAGCTTCTGCTGTGGCGCTTACACAGCTGGGGCAT 808
QY 351 GGGCTTCATCGCTTGGGGAGCGGGTGGCGGGTGAACGGGAAGTCCGAGGGTCTCT 410
DB 809 TGGCTTATGCTTGGCCAGGCAACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 868
QY 411 CTCCCGCTTTTGGCC--GGAAAGCCCCAAGAGAGAGCTTTTGGCCCTTGGCCGAG 468
DB 869 GTGCCGTGTCCGAGCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 928
QY 469 GGGCTCTCTCCGAGGGCGTGAACCGG-----GGGTGTGAACACAGCCCTCAT 518
DB 929 GGGTCTAGCCCAAGCTGTGTGAACCCAGCCCGGCGGAGATTTCAACCAAGACAGCAT 988
QY 519 GGAAGCTGGAGGCGAGGTGTGCTGTGCGAAACGAGCCCGTTGCGGGGCTTGGCCCTTAGG 578
DB 989 GGAGCTAGGGGCGACAGTGTGTATCCCAAGCGCCCACTGTGTGACAGCTGTGTGTGTGT 1048
QY 579 GGGCTTCTGCGGGG 592
DB 1049 GAGCTGTGCGGG 1062
```

RESULT 6  
US-09-974-300-1583  
; Sequence 1583, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 10085, 500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: PatSeq for Windows Version 4.0  
; SEQ ID NO: 1583  
; LENGTH: 528  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(528)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-974-300-1583

Query Match 9.5%; Score 92.8; DB 9; Length 528;  
Best Local Similarity 53.4%; Pred. No. 8.5e-12;  
Matches 227; Conservative 0; Mismatches 183; Indels 15; Gaps 1;

```
QY 17 AAGCCCTCTCGCTGTATCGGAGAAACGCCGCCCTCTCCCTGCGGGGAGAGG 76
DB 43 AAGATTATTTCTGTATGAGCAAGAAACGGGATCTGCTGCGCTCAGACAGCG 102
QY 77 ACCCTTACCGGCTCTGTCTCCAGGTCTTGTGACAGACCGGGGTGAGAGGCC 136
DB 103 ACCCTTATAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 162
QY 137 TCCCTATTACCGCGCTTCTGAGCGCTTCTCCACCCCTGAAAGGCGCTGCGGGCTT 196
DB 163 TTCCTATTCAACAACTTATGAAAGATTCCGACCGTGTGAACCGCTGCTGAAACCG 222
QY 197 CCGTGAAGAGTCTTGAAGGTCTGCAAGGGGCGGGCTTACTACCGCGGCGGAAACC 256
DB 223 ACGAAGAAAGTCTTAAAGCCTGGAGAGACTGGGATCTATTCAAGGGTTGGAAT 282
QY 257 TCCACCGC-----CTGCCGGAAGCGTGAAGAGCTTCCCGGAGCTTCC 301
DB 283 TGCAAGAGCGCTGTAGAGGAAGTTTACGAACGATACGAGGTGTGTGTGTGTGTGTGTGT 342
QY 302 CGAGCTTGGGGGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 361
DB 343 AGGAATTCGAGAGCTTAAAGGCGTGGCCCTTATCAAGAGGCGGGGTGTGATGATCG 402
QY 362 CTTTGGGGAGCGGGGTGCGGGTGTGAGACGGGAAGTCTCGAAGGTCTTCCGCTCT 421
DB 403 CTTCGAATCAGCGGTGCTCCGCGGTGAGACGGAATGTATGCGGGTATGTATCGGATTC 462
QY 422 TCGCC 426
DB 463 TTTCC 467
```

RESULT 7  
US-10-437-963-1325  
; Sequence 1325, Application US/10437963  
; Publication No. US2004012343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua



```

; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 1325
; LENGTH: 5850
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_101202C.1
US-10-437-963-1325

```

```

Query Match          9.4%; Score 91.2; DB 17; Length 5850;
Best Local Similarity 54.4%; Pred. No. 1.2e-11;
Matches 211; Conservative 0; Mismatches 168; Indels 9; Gaps 1;

```

```

QY 65 GGGGGGAGAGACCTTACCGCGTCTGCTCCGAGGCTCTTTCAGACAGACCCCGG 124
DB 263 GCAGAGGAGAGAGCGCTACGCGGTGTGGGTGTGGAGTGTGCTGACGATACGCGGG 322
QY 125 TGAGACAGCGCCCTCCCTATTATACCGCGCTTTCAGACGCTTTCACCGCTGAGGCC 184
DB 323 TGCCCGTGTGTGTGACTACTACTCCCGGTGATGTGCGCCCTGCGCCACCGTGAAGGCC 382
QY 185 TGCCCGGCGCTTCTCTGAGAGAGTCTTGAAGGTCTGCAAGGGCGCGCTACTACCGGC 244
DB 383 TGCCCGTGTGCGACGAGAGAGAGGTGAACGAGATGTGGCGCGCTGCGGTATTAACCGGA 442
QY 245 GGGCGGAAACCTTCCACCGCGCTGCGCGGAGCGT-----GAGAGAGTTCCCGCGA 295
DB 443 GGGCTCGATTCTTCTGAGAGAGCAAAATTTGAAAAAGCGAGTTCCCTGCA 502
QY 296 GCTTCGCGAGCTTCGAGGCGCTTCTGCTGCTCGGCGCTTACACCGCGCGCGGTGCGCT 355
DB 503 CAGCATAGCACTCCGGAAGTTCGTGCAATGGGGAATTACAGCAGAGCGCATTCCTT 562
QY 356 CCATCGCCTTCGAGGAGCGGGTGGCGCGGTGAACGGGAACGTCCGAGGGTCTCTCC 415
DB 563 CCATAGCCTTCAATGAGGTGTCTCTGTGTGATGAAATGTCCGATTATCAGCAGGC 622
QY 416 GCCTCTTCCCGCGGAGAAAGCCCAAGA 443
DB 623 TTACGCTATTCTGATTAACCAAGA 650

```

```

RESULT 8
US-10-425-114-32132
; Sequence 32132, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53131)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 32132
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Zea mays

```

```

; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73265D07_FLI
US-10-425-114-32132

```

```

Query Match          9.1%; Score 89.2; DB 13; Length 1660;
Best Local Similarity 50.7%; Pred. No. 4.4e-11;
Matches 282; Conservative 0; Mismatches 253; Indels 21; Gaps 2;

```

```

QY 60 CTGGCGGGGAGAGAGACCTTACCGCGTCTGCTCTCGAGGCTCTTCTGAGAGAGC 119
DB 278 CGGACGAGAGAGAGAGCGGTACCGGTGTGGGTGTCCGAGTATCTGACAGAGAC 337
QY 120 CGGGGTGAGAGAGCGCTTCCCTATTACCGCGCTTCTGAGAGCTTTCACCGTGA 179
DB 338 GCGGTGCGCGGTGTGTGCTTACTACAGCGGTGATGCGCGGTGCGCCACCGTACG 357
QY 180 GGCCTGAGCGCGCTTCCCTGAGAGAGTCTTGAAGGTCTGAGAGGGGCGGCTACTA 239
DB 398 AGCCTGCGCGCGCTACGAGAGAGGTGAACGAGATGTGGCGGCTTGGCTACTA 457
QY 240 CCGGCGGGGAGAACACTCC-----ACCGCTGCGCCGAGAGCTGAGAGCTTCC 290
DB 458 CCGTGGGCTCGATTCTGCTGAGAGAGCAAGCATATGAAAGGGGTGTTTCC 517
QY 291 CCGAGCTTCCGAGCTTCGAGGCGCTTCTGCTGCTGCGGCTTACACCGCGCGGT 350
DB 518 TTGACAGCGCTTACCGCTTCTGAGTGTGCTGCTTGGGATTAACAGCTGAGCAAT 577
QY 351 GGCCTCATCGCTTCGAGGAGCGGTGAGCGCGGTGAACGGAGCTTCGAGAGGTCT 410
DB 578 CGCTTCAATTGCTTCAAGAGGTGTCCAGTTGTGATGAAATGTGATACAGCAT 637
QY 411 CTCCCGCTTCTTCCCGGAGAAACCCCAAGAGAA-----GAGCTTTCCG 458
DB 638 CAGCAGGCTTTACACCATCTGTACCAACCAAGAGATCTCAACAGTGAAGATTCTG 697
QY 459 CTTGCGCGAGGCTCTCCCGGAGGCGGTGAGACCGGGGGTGTGAAACGAGCGCTCAT 518
DB 698 GGACTGTGATGTAATGTTTACCGTTTGAACAGAGACTTCAACCAACGATGAT 757
QY 519 GAGCTCGGGGCGACGCTGCTGCGGAACGCGCCGTTGCGGGGCTGCCCCCTAGG 578
DB 758 GAGTTAGAGCAACATTATGTAGCAAGCAAAAGCTGTTCTCAATGCCAGTCTC 817
QY 579 GGCCTTCTCGCGGGG 594
DB 818 TAGCCACTGCCAAGCG 833

```

```

RESULT 9
US-10-216-817-31
; Sequence 31, Application US/10216817
; Publication No. US20030129619A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; FILE REFERENCE: 03495.0233-00000
; CURRENT APPLICATION NUMBER: US/10/216,817
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 912
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-216-817-31

```

```

Query Match      8.9%; Score 86.6; DB 15; Length 912;
Best Local Similarity 50.2%; Pred. No. 1.9e-10;
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

QY 22 CTCCTCGCTGTTACCGGAAAAACGCCGCCCTCCCTGCGGGG---GAGAAAGAC 78
DB 61 CTTCTCGGTTGGTATACAGATCCGACCGGGATCTGCCCTGGAGACCCGGTGTACG 120
QY 79 CCTTACCGGCTCTGTCGAGAGTCTTTCGACGACGACCCGGGTGACAGACCTTC 138
DB 121 CCGTGCAGATCTGTGTACGAGATTCATGCTGACGACGACCGCCCGCGGTGCTG 180
QY 139 CCTATTACCGCCGCTTTCTGAGCGCTTTCACCCCTGAAGGCCCTGGCCGCGCTTCC 198
DB 181 GCGATCTGCGCGACTGGGCTGGGGGTGCCACGCGCTGCGACCGGCCAGCCAGC 240
QY 199 CTGAAGAAGTCTTAAAGTCTTGACAGAGGGGCGGCTTACTACCGCGGGGGAACACTC 258
DB 241 ACCGCCGATGTATTACGCGCTGGAGGCAAGCTGGGCTATCCAGCGCAAGCCGCTTA 300
QY 259 CACCGCTGCGCCGAAAGCGTG-----GAGGAGCTTCCCGAGCTTGCC 303
DB 301 CACGAGTGGCGCACCGTATGCGCCCGGACCAATGACGTGTGCCGACGATATGAG 360
QY 304 GAGCTTGGGGGCTTCTGTGCTTGCGCTTACACCGCGGCGGCGGTGCTTCATGCC 363
DB 361 ATCTGTGTCACCTGCGCGGGGCTGCGGAGCTACACCGCGCGGGGCTGTGCTTGGCT 420
QY 364 TTCGGGAGCGGGTGGCGGGGTGAGCGGGAAGCTCCGAGAGGCTCTCCGCTT- 422
DB 421 TACCGCAAGCGGGTCCGGGTGTGACACCAATGTGCGCGCGGTGGCGCTTGGCT 480
QY 423 -----CGCCCGGAAAAGCCCAAGAGAGAGCTTTTCGCTCGCCCAAGGC 471
DB 481 CACGCGCGCCCGACGCGGTGGCGCATCGGTGCCGACGACGCGACGCTTGGCG 540
QY 472 CTCCTCCCGAGGCGGTGACACCGGGGGTGTGAAACAGGCGCTCATGAGCTCGGAGCC 531
DB 541 CTGTGTCGCAACCGCGAGACGCGCGCTGAATTTTCGCTCGCTGATGAGTTGGGTGGC 600
QY 532 ACGGTCTGCTGCGAAGCGCCCGTTCGCGGGGCGTCCCTTGAAGGGGCTTCTGCGCG 591
DB 601 ACGGTGTGCAACCGCGCGACACCCCGGTGTGAGTATGCCGCTGACGTGTCCGATGG 660
QY 592 GGAAGAAGGCGCCCGGGCGCTTACCCCGCGCCCAAGAAAGCGCGGAGAGAGAGCGC 651
DB 661 CGGCAATGCCGTTATCCGCGTCCGACGAGTCC---GCCGCGCGGGGGGAGGCTTACAC 717
QY 652 CTCGTGCGCTCGTCTCTCTCGGGCGG 678
DB 718 GGAACGACCGCCCAAGTCCCGGACGG 744

RESULT 10
US-10-216-817-21
; Sequence 21, Application US/10216817
; Publication No. US20030129619A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING M. TUBERCULOSIS
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; FILE REFERENCE: 03495.0233-00000
; CURRENT APPLICATION NUMBER: US/10/216,817
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1312

```

```

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; US-10-216-817-21

Query Match      8.9%; Score 86.6; DB 15; Length 1312;
Best Local Similarity 50.2%; Pred. No. 1.8e-10;
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

QY 22 CTCCTCGCTGTTACCGGAAAAACGCCGCCCTCCCTGCGGGG---GAGAAAGAC 78
DB 261 CTTCTCGCTTGGTATCAGGATCGACCGGAGTCTGCCCTGGACGAGCCGGTGTACG 320
QY 79 CCTTACCGGCTCTGTCGAGAGTCTTTCGACGACGACCCGGGTGACAGACCTTC 138
DB 321 CCGTGCAGATCTGTGTACGAGATTCATGCTGACGACGACCGCCCGCGGTGCTG 380
QY 139 CCTATTACCGCGCTTTCTGAGAGCTTTCACCCCTGAAGGCCCTGCGCGGCTTCC 198
DB 381 GCGATCTGCGCGACTGGGCTGGGGGTGCCACGCGCTGCGACCGGCCAGCCAGC 440
QY 199 CTGAAGAAGTCTTAAAGTCTTGACAGAGGGGCGGCTTACTACCGCGGGGGAACACTC 258
DB 441 ACCGCCGATGTATTACGCGCTGGAGGCAAGCTGGGCTATCCAGCGCAAGCGCTTA 500
QY 259 CACCGCTGCGCCGAAAGCGTG-----GAGGAGCTTCCCGAGCTTGCC 303
DB 501 CACGAGTGGCGCACCGTATGCGCCCGGACCAATGACGTGTGCCGACGATATGAG 560
QY 304 GAGCTTGGGGGCTTCTGTGCTTGCGGCTTACACCGCGGCGGCGGTGCTTCATGCC 363
DB 561 ATCTGTGTCACCTGCGCGGGGCTGCGGAGCTACACCGCGCGGGGCTGTGCTTGGCT 620
QY 364 TTCGGGAGCGGGTGGCGGGGTGAGCGGGAAGCTCCGAGAGGCTCTCCGCTT- 422
DB 621 TACCGCAAGCGGGTCCGGGTGTGACACCAATGTGCGCGCGGTGGCGCTTGGCT 680
QY 423 -----CGCCCGGAAAAGCCCAAGAGAGAGCTTTTCGCTCGCCCAAGGC 471
DB 681 CACGCGCGCGCGACGCGGGTGGCGCATCGGTGCCGCGGACGACGCTTGGCG 740
QY 472 CTCCTCCCGAGGCGGTGACACCGGGGGTGTGAAACAGGCGCTCATGAGCTCGGAGCC 531
DB 741 CTGTGTCGCAACCGCGAGACGCGCGCTGAATTTTCGCTCGCGCTGATGAGTTGGGTGGC 800
QY 532 ACGGTCTGCTGCGAAGCGCCCGTTCGCGGGGCGTCCCTTGAAGGGGCTTCTGCGCG 591
DB 801 ACGGTGTGCAACCGCGCGACACCCCGGTGTGAGTATGCCGCTGACGTGTCCGATGG 860
QY 592 GGAAGAAGGCGCCCGGGCGCTTACCCCGCGCCCAAGAAAGCGCGGAGAGAGAGCGC 651
DB 861 CGGCAATGCCGTTATCCGCGTCCGACGAGTCC---GCCGCGCGGGGGGAGGCTTACAC 917
QY 652 CTCGTGCGCTCGTCTCTCTCGGGCGG 678
DB 918 GGAACGACCGCCCAAGTCCCGGACGG 944

RESULT 11
US-10-437-963-35788
; Sequence 35788, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B

```

; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 35788  
 ; LENGTH: 1425  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_39677C.1  
 US-10-437-963-35788

Query Match 8.8%; Score 86.2; DB 17; Length 1425;  
 Best Local Similarity 55.7%; Pred. No. 2.3e-10;  
 Matches 191; Conservative 0; Mismatches 143; Indels 9; Gaps 1;

65 GGGGGGAGAGAGACCCCTTACCGGCTCTGCTCCGAGTCTCTTTCAGACAGACCCGG 124  
 266 GCGAGAGAGAGAGAGGCTTACGCGGTGCTGAGAGTGAATGCTGACAGACGCGGG 325  
 125 TGAGAGAGGCTTCCCTTATTAACGCGCTTTCTGAGCGCTTTCCACCTTGAGGCCC 184  
 326 TACCGGTGTGTGCTACTACTCCGCTGATGCGCCGCTGGCCACCGTGAGACGCC 385  
 185 TGCGCGGCTTCCCTTGAAGAGTCTTGAAGTCTGAGGCGGCGGCTTACTACCGGC 244  
 386 TCGCGCGCGCCAGCAGAGAGAGTGAACGATGTGGCGGCTCGGATTTACCGGA 445  
 245 GGGCGGAGACCTTCAACGCGCTGGCCCGAAGCT-----GAGAGAGCTTCCCGCA 295  
 446 GGGCTCAATTTCTCTGAGAGAGCAAAACAAATTTGTAAGGAGGAGTCTTCCGCA 505  
 296 GCTTCCGAGCTTGGGAGCTTCTGTCTCTGAGGCTTACACCGCGCGGCTGCTCT 355  
 506 CAGCATCAACACTCGTGAAGTCTGCAATGGGAGTTACACAGCAGAGCATTGCTT 565  
 356 CCATCGCTTGGGAGAGCGGTGCGCGCGCTGACCGGAACT 398  
 566 CCAATGCTTCAATGAGTGTCTCTGTGTGAGCGGAAATGT 608

## RESULT 12

US-10-158-844-12/c  
 ; Sequence 12, Application US/10158844  
 ; Publication No. US20040029118A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kunsch et al.  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 391  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-R  
 ; OPERATING SYSTEM: Windows 98  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/158,844  
 ; FILING DATE: 03-Jun-2002  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/961,527  
 ; FILING DATE: 1997-10-30  
 ; APPLICATION NUMBER: US 60/029,960  
 ; FILING DATE: 1996-10-31  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Hyman, Mark J.  
 ; REGISTRATION NUMBER: 46,789  
 ; REFERENCE/DOCKET NUMBER: PB340P1D1

; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 9909 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
 US-10-158-844-12

Query Match 8.4%; Score 82; DB 13; Length 9909;  
 Best Local Similarity 52.2%; Pred. No. 1.3e-09;  
 Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

22 CTCCTGCTGATACCGGAGAAACGCCGCCCTCCCTCGCGGGGAGAGACCTT 81  
 5808 CTTCTGCTGTGTATGATGAAACAAAGATTTCTTGGAGAGAAATTAATCTT 5749  
 82 TACCGGTCTGTCTCCGAGTCTCTTCTGACAGACCGGCTGAGACGCTCTCCC 141  
 5748 TATCAATCTGGGTATCTGAATCATGCTTCAAGCAGACAGGCTGATACATTCCT 5689  
 142 TATTACCGCGCTTCTGTGAGCGCTTCCACCTTGAAGCCTTGGCGGCTTCCCTG 201  
 5688 TACTAGAAAGATTTTGTGACTGTTCACACTGTGAAAGTCTGGCACTGGCCTGAG 5629  
 202 GAAGAGTCTTGTAGGTCTGCGAGGGCGGCGCTA-----CTACCGGCGG 246  
 5628 GAAAGTTTACTGAAACTTGGAGGCTTGGCTATTTATTTCTGAGTTCCAAATATGAG 5569  
 247 GCGGAACACTTCCACCGCTGCGCCGAGAGGCTTCCCGAGCTTCTGCGCAG 306  
 5568 GCTGCAACCAAGATATGATGACTTGTGTGGCAATTTCCAAATACCTATGAGGA 5509  
 307 CTTGCGGCTTCTGTCTGAGGCTTACACCGCGCGGCTGCTCATGCTTCT 366  
 5508 ATTTCAGCTTGAAGGATTTGACCTTACACAGAGACCAATTTCCAGTATGCTTT 5449  
 367 GGGAGGGGTGCGCGCTGACGGAAGCTCGGAGGCTCTCCGCTCTTCCG 424  
 5448 AACTTGCTTGAAGCAGCTGATGATGATGATGAGGCTTGGCTTGGCTGCTTT 5391

## RESULT 13

US-09-864-866-46  
 ; Sequence 46, Application US/09864866  
 ; Patent No. US20020127656A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lloyd, R. Stephen  
 ; APPLICANT: McCullough, Amanda K.  
 ; APPLICANT: Nguyen, Khoa  
 ; TITLE OF INVENTION: DNA REPAIR POLYPEPTIDES AND METHODS OF USE  
 ; FILE REFERENCE: 265,00170101  
 ; CURRENT APPLICATION NUMBER: US/09/864,866  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/206,279  
 ; PRIOR FILING DATE: 2000-05-23  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 46  
 ; LENGTH: 968  
 ; TYPE: DNA  
 ; ORGANISM: Micrococcus luteus  
 US-09-864-866-46

Query Match 8.3%; Score 81.2; DB 9; Length 968;  
 Best Local Similarity 47.2%; Pred. No. 3e-09;  
 Matches 300; Conservative 0; Mismatches 318; Indels 18; Gaps 1;

44 ACGCCGCGCTTCCCTGCGGCGGAGAGACCTTACCGGCTCTGTCTCGAGG 103  
 200 ACCGTACGCGCTGCGGAGCTTTCAGAGCGCGTCTGAGCTCTGTGCGCAGG 259  
 104 TCCTTTCAGACACCGCGGTGAGCAGGCGCTCCCTATTACCGCGCTTCTGAGC 163



QY 321 TGGTCTCGGGCCCTTAACCGCGGGGGGTGGCCCTTCATCGCCTTGGGGAGCGGCTGGC 380  
Db 507 CGGCGTAGGACGCAAGACGGCTTCGTGGTGTCTGGCAACGCTTCGGGCGGCCCGCAT 566  
QY 381 GGGGTGAGCGGGAACGTCCGGAGGGTCTTCGCCCTTTCGCCCGGAAAGCCCA 440  
Db 567 CACGTGACACCCATTTCCAGCGGCTGTCCGCCGTGGCAGTGACCGAGAGAGA 626  
QY 441 GGAGAGAGAGCTTTGCGCCCTCGCCCAAGGCTCTCCCGAGAGGCTGACCCGGGGT 500  
Db 627 TCCGACAGATCGAGCGGCGCTTCGCCGAGAGAGGAGTGAAGATGCT 686  
QY 501 GTGAAACAGGCCCTCATGAGCTCGGGGCGACGCTCTGCCGAAACGGCCCGTTG 560  
Db 687 CTGCAACATG--TGATCTTCCACGGCGCGCATCTGTCAAGCCCGCAAGCCGGCTG 743  
QY 561 CGGGGCTGCCCCCTTAGGGGCTTTCGCCGGGGAGAGGCCCGGGGCTAACCCGC 620  
Db 744 CGGGGCTGCCCCCATGCCCCGCTCTGCCCCGCTACGGGAGGGCGAGACCGACCCGA 803  
QY 621 GCCCAGAGCGCCGGGGAAGAGAGAGCGCTGTGCCCTCTGTC 667  
Db 804 GAAGGCGAAGAGCTCTGAGTACGAGAAAGGGGCTTCCCGGCC 850

Search completed: October 8, 2004, 06:39:54  
Job time : 591 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 4, 2004, 03:57:10 ; Search time 2393 Seconds

(without alignments)  
12166.998 Million cell updates/sec

Title: US-09-938-901-1

Perfect score: 975  
Sequence: 1 gtggaagccggcggaagc.....tagtcccccccggaagca 975Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlun:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estlun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rnd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	10.7	1014	12	BM561672
2	101.4	10.4	1056	12	BM561672
3	100.2	10.3	496	12	BM561672
4	99.4	10.2	969	13	BM561672

5	95.2	9.8	1009	12	BM925099
6	91.6	9.4	964	13 <th>BM686151</th>	BM686151
7	90.6	9.3	662	12 <th>BM1818417</th>	BM1818417
8	90.2	9.2	577	12 <th>BM177663</th>	BM177663
9	90	9.2	771	12 <th>BM1818560</th>	BM1818560
10	90	9.2	1014	13 <th>BM401779</th>	BM401779
11	90	9.2	1933	11 <th>AK077546</th>	AK077546
12	88.6	9.1	1031	12 <th>BM921263</th>	BM921263
13	87.4	9.0	900	12 <th>BM761186</th>	BM761186
14	85.8	8.8	982	13 <th>BM415111</th>	BM415111
15	84.4	8.7	695	28 <th>BM2892784</th>	BM2892784
16	84.2	8.6	563	29 <th>CG898797</th>	CG898797
17	83.2	8.5	571	29 <th>CG898763</th>	CG898763
18	82.6	8.5	999	13 <th>BM685480</th>	BM685480
19	82.4	8.5	1201	9 <th>AL546337</th>	AL546337
20	82.2	8.4	1072	12 <th>BM925133</th>	BM925133
21	81.2	8.3	898	13 <th>BM685409</th>	BM685409
22	81	8.3	935	29 <th>CNS006XK</th>	CNS006XK
23	80	8.2	1179	12 <th>BM549153</th>	BM549153
24	79.4	8.1	556	29 <th>CG898299</th>	CG898299
25	79.2	8.1	982	13 <th>BM415111</th>	BM415111
26	79	8.1	641	12 <th>BM224720</th>	BM224720
27	78.4	8.0	360	12 <th>BM1343166</th>	BM1343166
28	78.4	8.0	530	10 <th>BM075702</th>	BM075702
29	78.4	8.0	532	10 <th>BM030801</th>	BM030801
30	78.4	8.0	551	10 <th>AM656251</th>	AM656251
31	78.2	8.0	1203	29 <th>CNS015Y4</th>	CNS015Y4
32	77.8	8.0	925	29 <th>CNS0091P</th>	CNS0091P
33	77.6	8.0	792	12 <th>BM1656309</th>	BM1656309
34	77.6	8.0	932	29 <th>CNS0072Q</th>	CNS0072Q
35	77.2	7.9	786	14 <th>CK021799</th>	CK021799
36	77.2	7.9	1203	29 <th>CNS015Y4</th>	CNS015Y4
37	76.6	7.9	828	13 <th>BM425797</th>	BM425797
38	76	7.8	1083	14 <th>CK206171</th>	CK206171
39	75.4	7.7	1202	28 <th>BM2051682</th>	BM2051682
40	75.4	7.7	769	12 <th>BM918941</th>	BM918941
41	74.6	7.7	865	13 <th>BM0890365</th>	BM0890365
42	74.4	7.6	563	12 <th>BM830888</th>	BM830888
43	74.4	7.6	580	12 <th>BM759737</th>	BM759737
44	74.4	7.6	594	12 <th>BM745229</th>	BM745229
45	74.4	7.6	981	12 <th>BM926983</th>	BM926983

## ALIGNMENTS

RESULT 1  
LOCUS BM561672  
DEFINITION AGENCOURT\_6593278 NIH\_MGC\_106 Homo sapiens CDNA clone IMAGE:5484344  
ACCESSION BM561672  
VERSION BM561672.1 GI:18807115  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1014)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-romail.nih.gov  
Tissue Procurement: Dr. Daniel McVicar, DMS/NCI  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LNCM2013 row: h column: 09  
High quality sequence stop: 617.

FEATURES  
source

Location/Qualifiers  
1.1014  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:548434"  
/tissue\_type="natural killer cells, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/lab\_lib="NIH\_MGC\_106"  
/note="Organ: Blood; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 10.7%; Score 104; DB 12; Length 1014;  
Best Local Similarity 54.0%; Pred. No. 5.6e-06;  
Matches 239; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

69 GGAGAAGACCTTACCGGCTCTGCTCCGAGTCTTTCGACAGACCCGGGTGA 128  
15 GGACAGCGGGCATATGCTGTGTGCTCAGAGTCATGTCAGACAGCCAGATTGC 74  
129 GCAGGCCCTCCCTATTACCGCGCTTTTCGAGCGCTTTCCACCCCTGAAGCCCTGC 188  
75 CACTGTGATCACTACTATACCGATGATGACAGAAAGTGGCCCTACCTCAGAGACCTGGC 134  
189 CGCGGCTTCCCTGGAAGAGGCTCTTAGGGTCTGCGAGGGGGGGGCTACTACCGGGGGGC 248  
135 CAGTCTTCCCTGAGAGAGGTGATCACTTGGCTGGCTGGCTGCTACTATTCTGTGG 194  
249 GGAACACTTCACCGCTGCGCCGGAAGCGTGAAGAGCTTCCCGGAGCTTC----- 300  
195 CGGGCGGTGACAGAGAGGAGCTCGAAGGTGATGAGAGGAGGAGGCGCATGTCACAG 254  
301 -----GCCAGCTTCCGGGGGCTTCTGTCTCCGGGCTTACACCGCGCGGT 350  
255 TACAGCAGAGACCTTCAGACAGCTCTGCTGGCGGTGAGGCGGTACACAGCTGGGGCCAT 314  
351 GGCCTCATCGCTTGGGGAGCGGGTGGCGGGGCTGGA CGGGAACGTCCGAGGGTCT 410  
315 TGCCTTATCGCTTTGGCCAGGCAACCGGTGTGTGATGGCAAGTTCACAGGATGCT 374  
411 CTCGCCCTCTTTCGCC--GGAAAGCCCAAGAGAGAGAGCTTTCCCTCGCCAG 468  
375 GTGCCGTGTCCGAGCATTTGTGTGATCCAGACACCTTGTTCACAGAGCTTG 434  
469 GGCCTCTCCCGAGGGCGGTGACCG-----GGGTGTGGAACCAAGCCCTCAT 518  
435 GGGTCTAGCCCAAGAGCTGTGAGCCAGCCCGGCGCAAGATTCACCAAGCAAGCCAT 494  
519 GGAGCTCGGGGCAAGGCTGTGCTGCGCAAGAGCGCCCTTACGGGGCTGCCCCCTAG 578  
495 GGAAGCTAGGGGCCAAGTGTGATCCCAAGCGCCCACTGTGACCAAGTCCCTGTGA 554  
579 GGCCTTTCGCGGG 592  
555 GAGCCTGTGCGGG 568

RESULT 2  
BF974701 1056 bp mRNA linear EST 22-JAN-2001  
LOCUS 602245343F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4336498 5',  
DEFINITION mRNA sequence.  
ACCESSION BF974701  
VERSION BF974701.1 GI:12341916  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1056)  
AUTHORS  
NIH-MGC <http://mhc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:  
<http://image.llnl.gov>  
Plate: LHCMI209 row: e column: 11  
High quality sequence stop: 678.

FEATURES

source

Location/Qualifiers  
1.1056  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4336498"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 10.4%; Score 101.4; DB 10; Length 1056;  
Best Local Similarity 53.7%; Pred. No. 1.4e-05;  
Matches 299; Conservative 0; Mismatches 226; Indels 32; Gaps 3;

69 GAGAAGACCTTACCGGCTCTGCTCCGAGGCTCTTCTGACAGACCCGGGTGA 128  
228 GGACAGCGGGCATATGCTGTGTGCTCAGAGTCTATGTCAGACAGCCAGGTTGC 287  
129 GCAGGCCCTCCCTATTACCGCGCTTTTCGAGCGCTTTCCACCCCTGAAGCCCTGC 188  
288 CACTGTGATCACTACTATACCGATGATGACAGAAAGTGGCTACCTCAGAGACTGGC 347  
189 CGCGGCTTCCCTGGAAGAGTCTTAGGCTGTGCAAGGGGGGGGCTACTACCGCGGGC 248  
348 CAGTCTTCCCTGAGAGGTGAAATCACTTGGGCTGGCTGAGCTACTATTCTGTGG 407  
249 GGAACACTTCACCGGCTGCGCCGAAAGCTGAGAGAGCTTCCCGAGCTTC----- 300  
408 CGGGCGGTGAGAGAGAGCTCGAAGGTGTGAGAGACTGAGGGGCCAATGCCAG 467  
301 -----GCCAGCTTCCGGGGCTTCTGTGCTCGGAGCTTACACCGCGGGCGG 349  
468 TACAGCAGAGACCTTTCAGAGACTCTGCTGGCGGTGGGCGCTACACAGTGGGCCA 527  
350 TGGCTCATAGGCTTTGGGAGCGGGTGGCGGGTGAACGGGAACGTCCGAGGGTCC 409  
528 TTGCTCTATGCTTTTGGCCAGCAACGGTGTGTGATGGCAAGTTCAGACGGGTGC 587  
410 TCTCCGCTCTTTCGCCC--GGAAAGCCCAAGAGAGAGCTTTCCGCTCGGCCA 467  
588 TGTGCTGTGTCGAGCATTTGTGTGATCCAGACACACTTGTTCACAGAGACTCTG 647  
468 GGCCTCTCTCCCGAGGGCGGTGACCC-----GGGGGTGTGGAACCAAGCCCTC 516  
648 GGGTCTAGCCAGCAGCTGTGTGAGCCAGCCCGGCAAGAGATTTCAACCAAGCAGAC 707



QY 517 ATGAGCTCGGGGCGGAGCTGCTGCTCCGAAAGCCCGCTTGCAGGCTGCCCCCTA 576  
 DB 708 ATGAGCTTACGAGGCGGCGGAGCTGCTCCGAAAGCCCGCTTGCAGGCTGCCCCCTA 767  
 QY 577 GGGGCTTTCGCGGGG 593  
 DB 768 TGGAGAGCTTGTGCGGG 784

RESULT 3  
 BG383035 496 bp mRNA linear EST 12-MAR-2001  
 LOCUS 3009920 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.  
 ACCESSION BG383035  
 VERSION BG383035.1 GI:13307507  
 KEYWORDS EST.  
 SOURCE Sus scrofa (pig)  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 1 (bases 1 to 496)  
 Fahnenkrug, S.C., Smith, T.P.L., Freking, B.A., Cho, J., White, J., Vallet, J., Wise, T., Rohrer, G.A., Pertea, G., Sultana, R., Quackenbush, J. and Keele, J.W.  
 Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly  
 Mamm. Genome 13 (8), 475-478 (2002)  
 12213789  
 12226715

TITLE Contact: Smith TPL  
 JOURNAL USDA, ARS, US Meat Animal Research Center  
 MEDLINE PO Box 166, Clay Center, NE 68933-0166, USA  
 PUBMED Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@meat.mars.usda.gov  
 Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -mismatch 12 options.  
 PCR primers  
 FORWARD: AGGATCAGCATGATGACCAT  
 BACKWARD: GTTATCCAGTCAGCAGC  
 Plate: 6 row: M column: 13  
 Seq primer: ATTTAGTGACATTAAG.  
 Location/Qualifiers  
 1..496  
 /organism="Sus scrofa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9823"  
 /rname\_type="pooled"  
 /lab\_host="DH10B"  
 /clone\_id="MARC 1P1G"  
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; library made from pooled tissue from day 11, 15, 20, and 30 embryos."

FEATURES  
 source  
 1..496  
 /organism="Sus scrofa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9823"  
 /rname\_type="pooled"  
 /lab\_host="DH10B"  
 /clone\_id="MARC 1P1G"  
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; library made from pooled tissue from day 11, 15, 20, and 30 embryos."

ORIGIN  
 Query Match 10.3%; Score 100.2; DB 12; Length 496;  
 Best Local Similarity 57.1%; Pred. No. 1.8e-05;  
 Matches 214; Conservative 0; Mismatches 143; Indels 18; Gaps 1;

QY 250 GAACACTCCAGCCGCTGCGCCCGAAGCGTGAAGAGCTTCCCGAGCTC----- 300  
 DB 231 CGCTGCTTACAGACAGAGCGCCCGGAAGCTGTGAGAGAGCTAGAGGCGACATCCAGCA 290  
 QY 301 -----GCCGAGCTTGGGGGCTTCTGTGCTCGGGCTTACACCGCGCGCGCTG 351  
 DB 291 ACAGACAGAGACCTCGACGCGCTCTGCGCGTGGCGGTACAGACGCGGGCCATT 350  
 QY 352 GCTTCATCGCTTTCGGGAGACCGGTGCGCGCTGAGACGGAAAGCTCCGAGGGCTTC 411  
 DB 351 GCTTCATTCCTTTTGGCAGGAGCTGTGTGTGAGAGGAAATGATTTGGGTGCTG 410  
 QY 412 TCCGCGCTTTCGCC 426  
 DB 411 TCCGCGCTTCGAGGC 425

RESULT 4  
 BX367285 969 bp mRNA linear EST 08-MAY-2003  
 LOCUS BX367285 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 DEFINITION Homo sapiens cDNA clone CS0DL004YB24 5'-PRIME, mRNA sequence.  
 ACCESSION BX367285  
 VERSION BX367285.1 GI:30449616  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 969)  
 Li, W.B., Gruber, C., Jesse, J. and Polayres, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by life Technologies, a division of invitrogen. This sequence belongs to sequence cluster 922.r For more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DL004YB24  
 Ping liang Email: liang@lifetech.com URL:  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID: CS0DL004YB24P1.

FEATURES  
 source  
 1..969  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL004YB24"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
 Query Match 10.2%; Score 99.4; DB 13; Length 969;  
 Best Local Similarity 53.1%; Pred. No. 2.7e-05;  
 Matches 296; Conservative 2; Mismatches 228; Indels 31; Gaps 3;

QY 189 CGCGGCTTCCCTGGAAAGAGCTCTTAGAGCTCTGGACAGGGGGCGGGCTACTACCGCGGGC 248  
Db 521 CAGTGCCTTCCCTGGAGAGAGTGAATCACTTGGGCTGGCGCTGAGGCTACTATTCTCGTGG 580  
QY 249 GGAAACACTTCACCGCGCTGGCCCCGAAAGCGTAGAGAGACTTTCGCCGAGCTTC----- 300  
Db 581 CCGGCGGCTGACAGAGGAGGAGCTCGAAAGAGTGTGAAGAGACTTAGGGGGCCACATGCCAG 640  
QY 301 -----GCCAGCTTGGGGGCTTCTGTCTCGGGCCTTACACCGGCGGGGAGT 350  
Db 641 TACAGCAGAGACCTTGACAGAGCTCTGTGCTGGCGTGGGCGCTCAACAGCTGGGGGCAT 700  
QY 351 GGCTTCATGCGCTTTCGGGAGCGGGTGGCGGGGAGTGAACGGGAAACGTCCGAGAGGCTT 410  
Db 701 TGCCTTATGCGCTTTTGGCCAGGCAACGCTGTGTGATGGCAACGTAGCAGCGGAGCTT 760  
QY 411 CTCCGCGCTCTTGGCCCC--GGGAAAGCCCAAGAGAAAGAGCTTTTGCCTCGCCCAAG 468  
Db 761 GTGCGCGGTCCGAGCATTGTGTCTGATCCAGCAGACCTTGTTTCCAGCAGGCTGTG 820  
QY 469 GGCTCTCTCCCGAGGGGCGTGAACC-----GGGGGTGTGAACCAAGGCCCTCA 517  
Db 821 GGGCTTAGCCCAAGCTGTGTGAACCAAGCCCGGCGCAAGAGATTCAACCAAGAGCCA 880  
QY 518 TGAAGCTCGGGGCGCACAGGCTTGCCTGGCGAAACGGCCCCCTTGGCGGGGCTCGCCCCTAG 577  
Db 881 TGAAGCTAGGGGCGCACAGTGTGAACCCCAAGCGCCCACTGTGACGCAATKCTCTGTGG 940  
QY 578 GGCGCTTTCGCGCGGGGG 594  
Db 941 AGAGCTGTTCCGGGAG 957

FEATURES	REFERENCE	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
RESULTS 5	BM925099	1009 bp	mRNA	linear	EST 12-MAR-2002											
BM925099	AGENCOURT.6627519 NIH_MGC_122 Homo sapiens	CDNA clone IMAGE:5762392	5', mRNA sequence.	BM925099	BM925099.1	GI:19375478	EST.	Homo sapiens (human)								
								Homo sapiens								
								Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.								
								1 (bases 1 to 1009)								
								NIH-MGC <a href="http://mgs.nci.nih.gov/">http://mgs.nci.nih.gov/</a> .								
								National Institutes of Health, Mammalian Gene Collection (MGC)								
								Unpublished (1999)								
								Contact: Robert Strusberg, Ph.D.								
								Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a>								
								Tissue Procurement: Life Technologies, Inc.								
								CDNA Library Preparation: Life Technologies, Inc.								
								DNA Sequencing by: Agencourt Bioscience Corporation								
								Clone distribution: MGC clone distribution information can be								
								found through the I.M.A.G.E. Consortium/LNL at:								
								<a href="http://image.lnl.gov">http://image.lnl.gov</a>								
								Plate: L1AM12612 row: 1 column: 17								
								High quality sequence stop: 655.								
								Location/Qualifiers								

```

FEATURES
SOURCE
Location/Qualifiers
1. 1009
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5762392"
/lab_host="DH10B"
/clone_11b="NH_MGC_122"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source:
anonymous pool of 34 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed

```

upon cloning. Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note this is a NIH\_MGC Library."

Query Match	9.8%	Score 95.2;	DB 12;	Length 1009;
Best Local Similarity	53.3%	Pred. No. 0.00012;		
Matches 289; Conservative	0;	Mismatches 223;	Indels 30;	Gaps 3;

QY	69	GGAGAAAGACCCCTTAAACCGCGCTCCGTGGCTCCGAGAGTCTCTTCTGACAGACACCCGGGTGA	128
Db	320	GGACAGCGGGGCAATTGCTGTGTGGGTCTCAGAGGTCATGCTCAGACAGACCCAGGTTGC	379
QY	129	GCAGGCGCTTCCCTTATTACCGCGCGCTTCTTGAGAGCTTCTCCACCGTGAAGGCCCTGGC	188
Db	380	CACGTGATCAACTACTATATCCGAGATGATGACGAAGTGGCTTACACTGACGAGACCTGGC	439
QY	189	CGGGCTTCCCTTGAAAGAGTCTTAGGGTCTTGACGAGGGGCGGGCTACTACCGAGGAGC	248
Db	440	CACTGCTTCCCTGAGAGAGGTGAATCACTGGGCTGGCTGGCTACTATTCTGTGG	499
QY	249	GGAACACTTCCACCGCGCTGGCCCGAGCGGTGAGAGACTTCCCCGAGCTTC-----	300
Db	500	CCGGCGGCTGCAAGAGAGAGCTGGAAGGTGTGAGAGACTAGGGGGCCAAATGCCACG	559
QY	301	-----GCCGAGCTTCGGGGGGCTTCTGTCTCGAGGCTTACACCGCGCGCGGT	350
Db	560	TACAGACGAGACCTGTCGACACACTCTGCTGTCGGCGGTACACAGCTGGGGCCAT	619
QY	351	GGCTTCATCGCTTTCGGGAGCGGGGTGCGGCGGTGACGAGAAAGTTCGAGAGGCTCT	410
Db	620	TGCTCTTATCGCTTTGGCCAGGCAACCGGTGTGTGATGTCACATGTAGCACGGGTGT	679
QY	411	CTCCGCGCTTTCGGCCC--GGGAAAGCCCAAGAGAAAGAGACTTTTGCCCTCGCCAG	468
Db	680	GTGCGGTGTCCGAGCCATTTGGTGTGATCCAGAGCAACCTTGTTTCCACAGACTCTGT	739
QY	469	GGCTCTCTCCCGAGGCGGTGACCCG-----GGGGTGTGAACAGAGCCCTCAT	518
Db	740	GGGTCTAGCCCGAGCGTGTGTGAACCGACCCGCGCAAGAAATTTCACCAAGACGCAT	799
QY	519	GGAGTTCGGGGCACGCTGTGCTGCGCAAAACGCCCCGTTCGGGGCTGCCCCCTAG	578
Db	800	GGAGGTAGGGGCAACAGTGTATACCCCAACGCCCCCACTGTGCAGCCAGTGGCCCTGTG	859
QY	579	AG	580
Db	860	AG	861

RESULT 6  
 B0686151  
 LOCUS  
 DEFINITION B0686151 964 bp mRNA linear EST 15-JUN-2002  
 5', mRNA sequence.  
 ACCESSION B0686151  
 VERSION B0686151.1 GI:21811455  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 1 (Mammalia; Eutheria); Primates; Catarrhini; Homnidae; Homo.  
 NIH-MGC http://mgc.ncl.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph. D.  
 Email: cgaabs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

### FEATURES

#### Location/Qualifiers

```
/mol_type="mRNA"
```

/db\_xref="taxon:9606"

```
/clone="IMAGE:6247891
```

```
/cissue_type="ductal carcinoma, cell line  
/job_host="DUN10B /rh3co-rootstart")"
```

```
/lad_novc=... (package
/cjone 1!b="NTH MSC 110"
```

```

/notes="Organ: pancreas:

```

Site 2: ECORI; cDNA made by oligo-dT priming.

**ORIGIN**

Query Match	Score	DB	Length
9.48	91.6	13	964

Qy	69	GGAGAAGGACCCCTTACCGCGTCCCTGGTCTCCGAGGCTCCCTTGCAGACACACCGGGTGA	128
Db	336	GGACAGGGGGGCATATGCTGTGTGTGGTCTCAAGAGTCATGCTGCAGACAACCCAGGTTGC	395
Qy	129	GCAGGCCCTCCCTATTATCCGCGCTTCTGGAGGCTTTCACCCCTAAGGCCCTGGC	188
Db	396	CACGTGTATCACTACTATACCGATGATGCAGAAAGTGGCCTTACACTGCAGAGCCTGGC	455
Qy	189	CGCGGCTTCCCTGGAAGAGTCTTAAAGGCTGCGCAGGGGGGGGCTACTACCGCGGCG	248
Db	456	CAGTGTCTCCCTGGAGGAGTGAATCACTCGGGCTGGCTGGGCTACTATTCTGTGG	515
Qy	249	GGAAACACTCCACCGCCTGCGCCCGAAGGTGGAGAGCTTCCCCGAGCTTC-----	300
Db	516	CCGCGCGCTGCAGAGGGAGGCTCGGAAGTGGTAAAGAGCTAAGGGGGGCACATGCGACG	575
Qy	301	-----GCCAGCTTCCGGGGGCTTCTGTGCTCCGGGCTTACACCGCGCGCGGCT	350
Db	576	TACAGCAGAGACCCCTGCAGCAGCTCTGTGTCGTGGGCGCTACACGCTGGGGGCAT	635
Qy	351	GGCCTCCATCGACCTTCGGGGAGCGGGTGGCGCGGCTGACGGGAACGTCCGAGGGGTCT	410
Db	636	TGCCTCTATCGCCTTTTGGCCAGGACACCGGTGTGTGTGATGGCAACGTAGACAGGGTGT	695
Qy	411	CTCCCGCCTTTCGCC--GGGAAAGCCCAAGAGAAAGACTTTTCGCTCGCCAG	468
Db	696	GTGCGGTGTCCGAGGCATTGGTGTGATCCAGCAGACCTTGTTTTCCAGAGGTGTG	755
Qy	469	GGCCTCTCTCCCGAAGGCGTGAACCG-----GGGTGTGGAACCAAGGCTTAT	518
Db	756	GGGTCTAAGCCAGCAGCTGTGTGAACCAAGCCCGGCAGAGAGATTTCACCAAGCAGCCAT	815
Qy	519	GGACCTCGGGGCGACGCTCTGCTGTCCGAAGAGGCGCGCTTTCGGGGGCTGCGCCCTAG	578
Db	816	GGACCTAAGGGGACACGTGTGTATCCCAAGGCGCCACTGTACAGCACTGGCGCTGGGG	875

RESULT 7	662 bp	mRNA	linear	EST 04-OCT-200
BI818417				
LOCUS				
DEFINITION	60303288491 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5174061 5' ,			
	mRNA sequence.			
ACCESSION	BI818417			
VERSION	BI818417.1 GI:15929249			

KEYWORDS	EST.
SOURCE	Homo sapiens (human

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (pages 1 to 662)

INTERNAT. JOURNAL. Published (1999)

**COMMENT**      **Contact:** Robert Strausberg, Ph.D

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Issue Procurement: Life Technologies, Inc.

CDNA Library Arrayed by: The I M A G E Consortium (I.M.I.)

DNA Sequencing by: Invitae Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Co

## FEATURES

### Location/Qualifiers

source . 1. .662

```
/organism="Homo sapien
```

```
/mol_type="miRNA"  
/db_xref="taxon:9606"
```

```
/u0_xlcl= cac011:3000  
/c]one="IMAGE:5174061"
```

```
/lab.host="DH10B"
```

```
/clone lib="NIH MGC 115".
```

ORIGIN

Query Match	Score	DB	Length
9.38	90.6	12	662

```

QY      69  GGAAGAAGACCCCTTACCGCGCTCGGTCTCCGAGTCTCTTCGACGAGAACCCGGGTGGA  128
Db      252  GGAAGAGCGGGCATATCTGTGTGTGGGTCTCAGAGSTCATGTGCAGCAGACCAAGTTGC  311
QY      129  GCAGGCCCTCCCTTATTACCGCGCTTCTGAGCGCTTTCACCCCTGAAGGCCCTTGC  188
Db      312  CACTGTGATCACTACTATATACCGGATGGATGACAGAAATGGCGCTACCTGCAGACCTGGC  371
QY      189  GCGGGCTTCCCTGGAGAAGTCTCTTAGGTTCTGGCAGGGGGCGGGCTATCTACGGCGGCG  248
Db      372  CAGTGTCTCCCTGGAGGAGGTGAATCAACTCTGGGCTGGCCCTGAGCTATCTATCTGTGG  431
QY      249  GGAACACTCCACCGCGCTGGCCCCGGAAGCTGTGAGAGACTTCCCCGAGCTTC-----  300
Db      432  CCGCGCGCTGACGAGAGGAGACTTCGGAAGTGTGTAGAGAGCTAAGGGGCCACATGCCAGC  491
QY      301  -----GCCGAGCTTCGGGGGCTTCTGTGCTTCGGGCTTACACCGCGCGAGCGGT  350
Db      492  TACAGCAGAGACCTGTGACGACGCTCTGTGGGTGGGGGCTACACAGCTGGGGCCAT  551
QY      351  GGCTTCATGCTCTTCGGGAGCGGGTGGCGGCGCTGAGCGGAACGTCCGAGAGGTCTT  410
Db      552  TGCCTCATATGCTCTTGTGANCAGGCACACGGTGTGTGTGATGGCAACGTAGCAGCGGTCT  611
QY      411  CTCGCCGCTCTGTGCC  426
Db      612  GTGCCGTGTCCGAGCC  627

```

RESULT 8  
BG177663 577 bp mRNA linear EST 06-FEB-2001  
LOCUS 602314663F1 NIH\_MGC\_85 Homo sapiens CDNA clone IMAGE:4420430 5',  
DEFINITION mRNA sequence.  
ACCESSION BG177663  
VERSION BG177663.1 GI:12684366  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgabs-remail.nih.gov  
Tissue Procurement: Louis Straud, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.jnl.gov  
Plate: LLM10157 row: n column: 15  
High quality sequence stop: 577.  
Location/Qualifiers

FEATURES  
source 1..577  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4420430"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lymph; Vector: PCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; cloned unidirectionally; oligo-dT primed.  
Average insert size 1.867 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 9.2%; Score 90; DB 12; Length 577;  
Best Local Similarity 55.3%; Pred. No. 0.00062;  
Matches 208; Conservative 0; Mismatches 150; Indels 18; Gaps 1;  
QY 69 GGAGAGACCTTACCGCGCTGCTCCGAGGTCTTTCGACAGACCCGGGTGA 128  
DB 186 GGACAGCGGGCATATGCTGTGTGAGGTCTGACAGGTCTGACAGACCCAGGTTGC 245  
QY 129 GCAGGCGCTCCCTATTACCGCGCTTCTGAGAGCGCTTCCACCCCTAAAGCCCTGAC 188  
DB 246 CACTGTGATCACTACTTATACCGATGATGACGAAGTGGCTTACACCTGACAGACCTTGC 305  
QY 189 CGCGGCTTCTGAGAGAGTCTTAGGGTCTGACAGGGGCGGCTACTACCGCGGAC 248  
DB 306 CAGTGTCTTCTGAGAGAGTGAATCACTTGGGCTGGCTGGCTACTATCTTCGTGG 365  
QY 249 GGAACACTCCACCGCTGCGCGGAAGCTGAGAGACTTCCCGAGCTTC----- 300  
DB 366 CCGGCGGTGACAGAGAGGCTCGAAGGTGTAGAGAGCTAAGGAGGCGCATGATCCACG 425  
QY 301 -----GCCGAGCTTGGGGGCTTCTGTCTCGGAGCTTACCGCGGGCGGCGGT 350  
DB 426 TACAGCAGAGACCTTGCAGACTCTGCTTGGGCGCTTACACAGCTGAGGCGCAT 485  
QY 351 GGCCTTCATCGCTTTCGGGAGCGGAGTGGCGGCTGACGCGAACTTCGAGGAGTCT 410  
DB 486 TGCCTTATCGCTTTCGGGCGAGCAACGAGTGTGATGCACTGACAGCGGCTCT 545  
QY 411 CTCGCGCTCTTGGC 426

Db 546 GTGCCGTGTCCGAGCC 561

RESULT 9  
BI818560 771 bp mRNA linear EST 04-OCT-2001  
LOCUS 603031084F1 NIH\_MGC\_115 Homo sapiens CDNA clone IMAGE:5174084 5',  
DEFINITION mRNA sequence.  
ACCESSION BI818560  
VERSION BI818560.1 GI:15929676  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgabs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.jnl.gov  
Plate: LLM11433 row: h column: 21  
High quality sequence stop: 757.  
Location/Qualifiers

FEATURES  
source 1..771  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5174084"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_115"  
/note="Organ: pooled brain, lung, testis; Vector:  
PCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
source anonymous pool of 6 male brains, age range 23-27; 1  
male lung, age 27; and 1 male testis, age 69. Library is  
oligo-dT primed and directionally cloned (EcoRV site is  
destroyed upon cloning). Average insert size 1.8 kb.  
insert size range 1-3 kb. Library is normalized and  
enriched for full-length clones and was constructed by C.  
Gruber (Invitrogen). Research Genetics tracking code  
021. Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 9.2%; Score 90; DB 12; Length 771;  
Best Local Similarity 55.3%; Pred. No. 0.00065;  
Matches 208; Conservative 0; Mismatches 150; Indels 18; Gaps 1;  
QY 69 GGAGAGACCTTACCGCGCTGCTCCGAGGTCTTTCGACAGACCCGGGTGA 128  
DB 253 GGACAGCGGGCATATGCTGTGTGAGGTCTGACAGGTCTGACAGACCCAGGTTGC 312  
QY 129 GCAGGCGCTCCCTATTACCGCGCTTCTGAGAGCGCTTTCGACAGGCGGCTTGC 188  
DB 313 CACTGTGATCACTACTTATACCGATGATGACGAAGTGGCTTACACCTGACAGACCTTGC 372  
QY 189 CGCGGCTTCTGAGAGAGTCTTAGGGTCTGACAGGGGCGGCTACTACCGGGGCG 248  
DB 373 CAGTGTCTTCTGAGAGAGTGAATCACTTGGGCTGGCTGGCTACTATCTTCGTGG 432  
QY 249 GGAACACTCCACCGCTGCGCGGAAGCTGAGAGCTTCCCGAGCTTC----- 300  
DB 433 CCGGCGGTGACAGAGAGGAGTGAAGTGTGAGAGAGCTAAGGGGCGCACATGCCACG 492  
QY 301 -----GCCGAGCTTGGGGGCTTCTGTCTTCGGGCTTACACCGGCGGCGGT 350  
DB 493 TACAGCAGAGACCTTGCAGACTCTGCTTGGCGGTGGGCGGCTACACAGCTGGGCGCAT 552

Qy 351 GGCCTCCATGCGCTTCGGGAGCGGCTGCGCGGAGAGGCTCT 410  
 Db 553 TGCCTCTATGCGCTTCGGGAGCGGCTGCGCGGAGAGGCTCT 612  
 Qy 411 CTCCTGCTGCTTCGGC 426  
 Db 613 GTGCGGCTGCTCCGAGC 628

RESULT 10  
 BX401779 1014 bp mRNA linear EST 13-MAY-2003  
 LOCUS BX401779 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 DEFINITION Homo sapiens cDNA clone CS0DL008BD01G02 5-PRIME, mRNA sequence.  
 ACCESSION BX401779  
 VERSION BX401779.1 GI:30630494  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1014)  
 L.I.W.B., Gruber, C., Jesse, J., and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 JOURNAL Contact: Genoscope  
 COMMENT Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 922.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DL008BD01Q1P1cluster=922.r. Contact :  
 Feng Liang Email : fliang@life.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DL008BD01Q1.  
 Location/Qualifiers  
 1..1014  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL008BD01"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /clone\_line="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
 Query Match 9.2%; Score 90; DB 13; Length 1014;  
 Best Local Similarity 55.3%; Pred. No. 0.00069;  
 Matches 208; Conservative 0; Mismatches 150; Indels 18; Gaps 1;

Qy 69 GGAAGAAGACCTTACCGCGCTGCTCCGAGGCTCTTGCAGACCGCGGCTGGA 128  
 Db 487 GGAAGCGCGGCAATGCTGTGTGCTGTGAGGTCTGAGAGCTGCTGAGACCGCGGCTGTC 546  
 Qy 129 GCAAGCCCTCCCTATTACCGCGCTTCTGAGCGCTTCCACCTGAAAGCCCTGGC 188  
 Db 547 CACTGTATCAACTACTATACCGGATGATCAGAAAGTGCGCTACACTGCAAGACCTGGC 606  
 Qy 189 CGCGGCTTCCCTGGAAGGCTCTTAAAGGCTGCGAGCGGCGGCTACTACCGCGGCGC 248  
 Db 607 CAGTCTTCTCTGAGAGGAGTGAATCAACTCTGCGGCTGCGCTGAGCTACTTCTCGTGG 666  
 Qy 249 GGAACACCTCCACCGCGCTGCGCGAAGCGTGAAGAGCTTCCCGGAGCTTC----- 300  
 Db 667 CCGCGGCTGAGAGGAGGAGCTCGGAAGGTGTAGAGAGCTAAGGCGGCACATGCCACG 726

Qy 301 -----GCCGAGCTTGGGGGCTTCCTGCTGCGGCTTACACCGGGGGCGGT 350  
 Db 727 TACAGCAGAGACCTTGCAGCAGCTTCCTGCGGCTGCGGCGCTACACAGCTGGGCGCAT 786  
 Qy 351 GGCCTCCATGCGCTTCGGGAGCGGCTGCGCGGAGAGAGGCTCT 410  
 Db 787 TGCCTCTATGCGCTTCGGGAGCGGCTGCGCGGAGAGAGGCTCT 846  
 Qy 411 CTCCTGCTGCTTCGGC 426  
 Db 847 GTGCGGCTGCTCCGAGC 862

RESULT 11  
 AK077546 1933 bp mRNA linear HTC 18-SEP-2003  
 LOCUS AK077546  
 DEFINITION Mus musculus 8 days embryo whole body cDNA, RIKEN full-length  
 enriched library, clone:573044C03 product:similar to ADENINE-DNA  
 GLYCOSYLASE [Mus musculus], full insert sequence.  
 ACCESSION AK077546  
 VERSION AK077546.1 GI:26097402  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Mech. Enzymol. 303, 19-44 (1999)  
 JOURNAL 99279253  
 MEDLINE 10349636  
 PUBMED 10349636  
 REFERENCE 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 JOURNAL 20499374  
 MEDLINE 11042159  
 PUBMED 11042159  
 REFERENCE 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, T., Nishi, K., Kitanaka, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
 Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kasai, Y.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuyama, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 JOURNAL 20530913  
 MEDLINE 11076861  
 PUBMED 11076861  
 REFERENCE 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 JOURNAL 11076861  
 MEDLINE 11076861  
 PUBMED 11076861  
 REFERENCE 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 JOURNAL 11076861  
 MEDLINE 11076861  
 PUBMED 11076861  
 REFERENCE  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
 Katoh, H., Kawai, J., Kondo, S., Konno, H., Kouda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohnishi, N.,

TITLE  
JOURNAL

Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

## COMMENT

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://phantom.gsc.riken.go.jp/.

## FEATURES

## source

1. 1933  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM DB:5730444C03"  
/db\_xref="MGI:2351728"  
/db\_xref="taxon:10090"  
/clone="5730444C03"  
/tissue\_type="whole body"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="8 days embryo"  
1. 1933  
/note="similar to ADENINE-DNA GLYCOSYLASE [Mus musculus] (SPTR/O99P21, evidence: FASTV, 83.8%ID, 100%length, match=1831)"

## ORIGIN

## Query Match

Best Local Similarity 51.1%; Score 90; DB 11; Length 1933;  
Matches 287; Conservative 0; Mismatches 245; Indels 30; Gaps 2;

```

QY 69 GGAAGAAGACCTTACCGCGCTGCTCCGAGCTCTTTCGACAGACCCGGGTGA 128
DB 323 GGAAGAAGAGGCGCTATGCTGTGGGTGTAAGGTTATGCTGACAGACCCAGGTTGC 382
QY 129 GCAAGCGCTCCCTATTATACCGCGCTTTCGAGCGCTTCCACCCCTGAAGCCCTGGC 188
DB 383 CACAGTGATGACCTATTATCTGATGATGAGAGTGCCAAAGCTTCAAGACCTGGC 442
QY 189 CGCGGCTTCCCTGGAAGAGTCTTAGGTTCGAGGGGGGGGCTACTACCGCGGAGC 248
DB 443 CAGCGCTTCCCTGGAAGAGTGAACAGGCTGCTGCTGCGCTGCTACTATTCGAGG 502
QY 249 GGAACACCTTCACCGCGCTGCGAGCGTGGAGAGCTTCCCGGAGCTTGC----- 301
DB 503 CCGTCGCTTACAAGAGAGCTAGAGAGTGTATAGAGAGCTAGAGAGCCATATGCCAG 562
QY 302 -----CGAGCTTCCGGGGCTTCTGCTGCTGCGGCTTACACCGGCGGCGGT 350
DB 563 TACAGAGAGACCTTACAGACAGCTCTCTGCTGCTGCTGCGGCGGTAACAGGCTGAGCAT 622
QY 351 GGCCTTCATGCGCTTTCGAGAGCGGAGTGGCGGCTGAGCGGAAGCTCCGAGAGTCT 410
DB 623 TCCCTTCATGCGCTTTCGAGAGCGGAGTGGCGGCTGAGCGGAAGCTTTCAGCGGTGCT 682
QY 411 CTCGCGGCTTTCGCGCGGAGAGCCCAAGAGAGAGAGCTTTCGCGCTGCGCGAGG 470
DB 683 GTGCGGTGTCGCGGCAATGTCGATCCACAGACAGCTTGTCTCATCAGCTG 742
QY 471 CTCCTCCCGAGGCGGTG-----ACCGGGGGGTGTGAGACCAAGGCGCTCAT 518

```

```

DB 743 GAATTAGCCACAGACTGCTGAGACCCAGCCCGGCTGGGACTTCAATCAAGCTGCCAT 802
QY 519 GGAAGCTCGGGGCCACCGCTGCTGCTGCGGAAACGCGCCGTTGCGGGGCTCCCTTGG 578
DB 803 GGAAGCTGGGGGCTACAGTGTGACCCCAAGGCGGCTTCTGACGACCTTCCCGGTGA 862
QY 579 GGCCTTCTCCCGGGGGAAGAG 600
DB 863 GAGCTTATGCGCGGCAATACAG 884

```

## RESULT 12

BM921263 1031 bp mRNA linear EST 12-MAR-2002  
LOCUS AGENCOURT\_6626344 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5752647  
DEFINITION 5', mRNA sequence.

## ACCESSION

BM921263.1 GI:19371642

## VERSION

BM921263

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## REFERENCE

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## AUTHORS

NIH-MGC http://mgc.nci.nih.gov/.

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

## COMMENT

Contact: Robert Strauberg, Ph.D.  
Email: gsapbs-remail.nih.gov

## Tissue Procurement

Life Technologies, Inc.

## cDNA Library Preparation

Life Technologies, Inc.

## DNA Sequencing

Arrayed by: The I.M.A.G.E. Consortium (LLNL)

## Clone distribution

MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

## http://image.llnl.gov

## plate: LLAM2787

High quality sequence stop: 655.

## Location/Qualifiers

1. 1031  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5752647"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_115"  
/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site: 1; Nori; Site 2; EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed as Gruber (in vitro). Research Genetics tracking code 021. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 9.1%; Score 88.6; DB 12; Length 1031;  
Best Local Similarity 53.8%; Pred. No. 0.0011;  
Matches 290; Conservative 0; Mismatches 219; Indels 30; Gaps 4;

```

QY 69 GGAAGAAGACCTTACCGCGCTGCTCCGAGCTCTTTCGACAGACCCGGGTGA 128
DB 321 GGAAGAAGAGGCGCAATGCTGTGGGTGTAAGGTTATGCTGACAGACCCAGGTTGC 380
QY 129 GCAAGCGCTCCCTATTATACCGCGCTTTCGAGCGCTTCCACCCCTGAAGCCCTGGC 188
DB 381 CACGTGATCAACTATACCTACCGATGAGTACGAAGTGGCCCTACACTCAGAGACTGGC 440
QY 189 CGCGGCTTCCCTGGAAGAGTCTTAGGTTCGAGGGGGGGGCTACTACCGCGGAGC 248
DB 441 CAGTGTCTCCCTGGAAGAGTGAATCAACTCTGGGCTGGCTGGGCTACTATTCTGTTG 500

```

QY 249 GGAACACCTCCACCGCCTTGCCCGAAGCGTGAAGAGCTTCCCGAGCTTC----- 300  
 Db 501 CCGGCGGCTGACGAGGAGAGCTTCGGAAGTGTGTAAGAGACTAGGGGGCCACATGCCACG 560  
 QY 301 -----GCCGAGCTTCGGGGGCTTCCTGTCTCGAGCCTTAACACCGCGCCGCGGT 350  
 Db 561 TACAGCAGAGACCTTCGACGAGCTCTGCTGCGTGGGGCGCTACACAGCTGGGGCCAT 620  
 QY 351 GGCTTCATCGCTTCCTGGGGA-GCGGCTGGCGGCGCTGGAAGGAACTCCGAGGGTCC 409  
 Db 621 TGCCTATATCCCTTTGGCCAGGCAACCGGTGTGTGATGGCAACCTAGACCGGGTGC 680  
 QY 410 TCTCCCGCTCTGGGCC---GGGAAAGCCCCAAGAGAGAGCTTTTCGCCCTCGGCC 466  
 Db 681 TGTGCTGTGTCCGAGCCATGTGTCTGATCCAGACAGACCTTGTTCACAGACCTCT 740  
 QY 467 AGGACCTCTCTCCCGAGGCGGTGAC-----CGGGGTGTGGAACCAAGCCCTCAT 518  
 Db 741 GGGGTCTAGCCACAGAGCTGTGTGACCAACCGGGCCAGGAATTCACCAAGAGCGCAT 800  
 QY 519 GGAGCTCGGGCCACGCTGTGCTGCGGAACGGCCCTTGGCGGGCTTGCCTCTAG 577  
 Db 801 GGAGCTAGGGGCCCCAGTGTGTACCCCAACGCCCACTGTGACAGCGGCGCTGTGG 859

RESULT 13  
 Bg761186 900 bp mRNA linear EST 15-MAY-2001  
 LOCUS 602718527F1 NIH\_MGC\_49 Homo sapiens cDNA clone IMAGE:4858374 5',  
 DEFINITION mRNA sequence.  
 Bg761186  
 VERSION Bg761186.1 GI:14071839  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 900)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@b-remail.nih.gov  
 Tissue Procurement: ATCC/DCTD/BTP  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: L1CM1712 row: F column: 07  
 High quality sequence stop: 867.

## FEATURES

source  
 1. 900  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4858374"  
 /tissue\_type="melanotic melanoma, high MDR (cell line)"  
 /lab\_host="RDH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_49"  
 /note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCAAGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 library."

ORIGIN

Query Match 9.0%; Score 87.4; DB 12; Length 900;  
 Best Local Similarity 54.9%; Pred. No. 0.0016;  
 Matches 207; Conservative 0; Mismatches 151; Indels 19; Gaps 1;  
 QY 69 GGAAGAGACCTTACCGCGCTCTGCTCCGAGGTCTTTCGACGAGACCCGGGTGA 128  
 Db 331 GACAGAGCGGCGCATATGCTGTGTGTCTCAGAGTCACTGTGACGAGACCCAGATTGC 390  
 QY 129 GCAGGCCCTCCCTATTATACCGCCGCTTCTGAGGCGCTTCCACCCTGAAGGCCCTGAC 188  
 Db 391 CACTGTGATCAACTATATACCGATGATGATCAGAAAGTGGCTTACACTGACAGACCTGAC 450  
 QY 189 GCGGCTTCCCTGGAAGAGTCTTAAAGGTCTGACGAGGGCGGCTACTACCGCGGAC 248  
 Db 451 CAGTCTTCCCTGAGAGAGTGAATCAACTTGGGCTGGCCGTACTATTCGTGG 510  
 QY 249 GGAACACTTCACCGCGCTGCGCCGGAAGCGTGAAGAGCTTCCCGAGCTTC----- 300  
 Db 511 CCGGCGGCTGACAGAGGAGGAGCTCGAAGGTGTAGAGAGCTAAGGGGCCACATGCCACG 570  
 QY 301 -----GCCGAGCTTGGGGGCTTCTGTGCTGGGCTTACACCGCGCGGGG 349  
 Db 571 TACAGCAGAGACCTTGCAGAGCTCTGCTGGCGTGGGCGCTACACAGCTGGGGCA 630  
 QY 350 TGGCTCCATGCTTCGGGAGCGGAGCGGCTGAGCGGGAACCTCCGAGGGTCC 409  
 Db 631 TTGCTCTATGCTCTTGGCCAGGCAACCGGTGTGTGTGATGCAACTAGCAGCGGTGC 690  
 QY 410 TCTCCCGCTCTTGGCC 426  
 Db 691 TGTGCGGTGTCGAGCC 707

RESULT 14  
 Bx415111/c 982 bp mRNA linear EST 15-MAY-2003  
 LOCUS Bx415111 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004Y120  
 DEFINITION Bx415111  
 ACCESSION Bx415111  
 VERSION Bx415111.1 GI:30765470  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 982)  
 Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. Contact : Feng Liang Email : fliang@life tech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Paradey Avenue Genoscope sequence ID : CS0CAP004B10Q71.

## FEATURES

source  
 1. 982  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0CAP004Y120"  
 /tissue\_type="THYMUS"  
 /clone\_lib="Homo sapiens THYMUS"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."

ORIGIN

Query Match 8.8%; Score 85.8; DB 13; Length 982;

```
Best Local Similarity 31.4%; Pred. No. 0.0029;
Matches 180; Conservative 119; Mismatches 274; Indels 0; Gaps 0;

QY 166 TTTCCACCTGAGACCCCTGCGCGGCTTCCCTGAGAGAGTCTTGAAGTCTGAGC 225
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 668 TTTTCTGSCSCTGGGCCCCCCCCCGSSGGTCCCCCTSMGTTCCTTGTGCGGGKOT 629

QY 226 GGGGCGGGCTACTACCGCGGGCGGAAACCTTCACCGCTTGGCCCCGAAAGCTGAGAG 285
Db 628 GGTGTGGGGGGGGTGGGTGGTGTGCTGCTCTCTCCCGCTGCTGCTGCTGCTGCTGCT 569

QY 286 CTTCCTCCCGAGCTTGGCCGACGTTGGGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 345
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 568 YTGSTGTGCTGTGTGGGGGCGCGGGGGGCGCGSSGGGCGCKGCGGGCGCGSSCGCC 509

QY 346 GCGGTGGCTCTCATGCGCTTTCGGGGGAGCGGGTGGCGGCGCTGAGACGGAACGTCCGAG 405
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 508 CCCCCCCCCCGSSGGCCCCCCCCCGCGGCCCCCGGGGCCCCCGCGCGCGCGCGCGCG 449

QY 406 GTCTCTCCGCTTTCGCGCGCGGAAAGCCCCCAAGAGAGAGCTTTTCCGCTTCCGCT 465
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 448 GCGGSSCCCGCBSCCGSSCGSSCGSSCGSSCGGGCGGGCGSSSSSKSCCGSSCGCGCC 389

QY 466 CAGGGCTCTCTCCCGAGCGCGGTGAGACCGGGGGGTGAGACAGGCGCTCATGAGACTC 525
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 388 CGSSSCCGCGSSGSSCGSSGSSCGCGSSGSSCGSSGSSGSSCGCGGCGCGCGCGCG 329

QY 526 GGGGCGACGCTGTGCTGCGCGAAACGCGCGCTTTCGGGGGCGCTGCGCGCGCGCTTC 585
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 328 SSSSSCGCGCGGSSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 269

QY 586 TCCCGGGGAGAGAGCGCGCGCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 268 GCGCGCGCGGGGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 209

QY 646 GAGCGCTCTGTGCGCGCTGCTCTCTCGCGGAGAGGGGTGACCTGAGAAAGCTTGAAG 705
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 208 CGCGCGGGGGGAGCGGGAGCGSSGGGGGCGSSBSGGSCCGCGCGRGGGGSSSSS 149

QY 706 GCGCGCTTTCAGAGGCTCTTACGCGCGCTTCCCGCTC 738
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 148 SSSSSSSSSSSBCBSSSSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 116

RESULT 15
BZ892784 695 bp DNA linear GSS 30-JUL-2003
LOCUS Hm8_0123_x1_020.ab1 Hm pUC18 Library Halocaula marismortui genomic
DEFINITION 5', genomic survey sequence.
ACCESSION BZ892784
VERSION BZ892784.1 GI:33343417
KEYWORDS GSS.
SOURCE Halocaula marismortui
ORGANISM Halocaula marismortui
          Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
          Halobacteriaceae; Halocaula.
REFERENCE 1 (bases 1 to 695)
AUTHORS Goo, Y., Roach, J., Glusman, G., Baliga, N.S., Deutech, K., Pan, M.,
          Dasgupta, S., Ng, W. V. and Hood, L.
TITLE Low-pass Sequencing for Microbial Comparative Genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Geo X
          Institute for Systems Biology
          1441 North 34th Street, Seattle, WA 98103, USA
          Tel: 206 732 1412
          Fax: 206 732 1299
          Email: ygoos@systemsbiology.org
          Seq primer: M13 forward
          Class: shotgun.
FEATURES
source 1..695
          Location/Qualifiers
          /organism="Halocaula marismortui"
          /mol_type="genomic DNA"
```

```
/strain="ATCC 43049"
/db_xref="taxon:2238"
/clone_lib="Hm pUC18 Library"
/notes="Vector: pUC18; Site_1: SmaI; A shotgun library was
constructed from Halocaula marismortui genomic DNA using
pUC18/SmaI/BAP plasmid"

ORIGIN
Query Match 8.7%; Score 84.4; DB 28; Length 695;
Best Local Similarity 51.2%; Pred. No. 0.0044;
Matches 292; Conservative 0; Mismatches 251; Indels 27; Gaps 3;

QY 75 GAACTTACCGCGCTCGGTCTCCGAGGTCTTCTGAGAGACCGGGTGAACAGAC 134
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 20 GGGCGGTACGAGATTTCGTTTCCGAGGTGATGACCAAGACCACTGACAGAGT 79

QY 135 CTTCCTATTATACCGCGCTTTCGAGCGGCTTTCACCGTGAAGGCGCTTGGCGCGCG 194
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 80 CGTGAAGCGCTGGAGAGATTTCGAGCGGTGGCGGACCGGAGCGGCGGCGGAGG 139

QY 195 TTCCCTGAGAGGTCTTGAAGGTGCG-----CAGGGGCGGGCTACTACCGCGGGC 248
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 140 CGACCGGTGAGCGGTGGGTCTTGGACCAAGTCACTGCTGGGATCAACCAACCGGCG 199

QY 249 GGAACACCTTCAACCGCGCTGGCGCGGAGCGTGAAG-----GAGCTTCCCGC 293
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 200 GAAGTATCTCCAGAGACGCGCGTCAAGTGTGATGATGATGATGATGATGATGATGATG 259

QY 294 GAGCTTTCGCGAGCTTTCGGGGGCTTCTGCTGCTGCGGCTTTCACCGCGCGCGGCTGCG 353
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 260 CGACCGGAGCGGCTTCAACGAGACTGATGAGGGGTGGTCCCTACACCGCAACGCTGCGC 319

QY 354 CTTCATTCGCTTTCGAGGAGCGGGTGGCGCGGTGAGCGGAACTTCGAGAGGTCTCTC 413
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 320 CTCTTTCGCTTTCACCAACGAGGAGCGGTGTGACGAGACGTAACGAGCGGCTCTGTA 379

QY 414 CGGCTTTCGCGCGGAGAGCGCGCAAGAGAGGCTTTTCGCTTTCGCGCGCGCGCT 473
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 380 CGGTGCTTTCGACGTG-----CTGACGACGATTCGGCGTTTGAAGCGCTGCGGCGAC 433

QY 474 CTTCGCGAGGCGGTGAGCGCGGGGTGTGAAACAGGCGCTGATGAGCTTGGGCGCAC 533
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 434 GCTCATGCGAGGAGGCGAGTGGCGGGCTTGGAAACAACCTATCATGGAATTGGGTGT 493

QY 534 GGTTCGCTTTCGAGAAAGCGCGCGCGCTTTCGCGGGCGCTTTCGCGGGCG 593
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 494 CGCTTTCGAGAAAGCGCGGATTTGAGCGGCGCACAGTGCCTTGGCGGAGGTGTTC 553

QY 594 GAAAGAGGCGCGCGCGCTACCGCGCGC 623
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 554 GGCATACGAGCGGCGGATTTTACCGCGAC 583

Search completed: March 4, 2004, 05:46:54
Job time : 2403 secs
```



Result No.	Score	Query Match	Length	DB	ID	Description
1	975	100.0	975	10	US-09-938-901-1	Sequence 1, Appl
2	112.4	11.5	939	14	US-10-156-761-4691	Sequence 4691, Ap
c	112.4	11.5	9025608	14	US-10-156-761-1	Sequence 1, Appl
	104	10.7	1878	9	US-09-925-301-484	Sequence 484, Appl
	92.8	9.1	528	9	US-09-974-300-1583	Sequence 1583, Ap
6	89.2	9.1	1660	12	US-10-425-114-2132	Sequence 2132, A
5	89.2	9.1	1660	12	US-10-425-114-2132	Sequence 2132, A
6	86.6	8.9	912	14	US-10-216-817-31	Sequence 31, Appl
7	86.6	8.9	1312	14	US-10-216-817-31	Sequence 31, Appl
8	86.6	8.9	1312	14	US-10-216-817-31	Sequence 31, Appl
9	82	8.4	9909	12	US-10-158-884-12	Sequence 12, Appl
c	82	8.3	968	9	US-09-864-866-46	Sequence 46, Appl
	81.2	7.3	930	14	US-10-156-761-4577	Sequence 4577, Ap
	71.4	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appl
12	71.4	7.0	991	15	US-10-338-221-1588	Sequence 1588, Ap
13	68.2	7.0	3155	15	US-10-338-221-1596	Sequence 3596, Ap
14	66.8	6.9	11058	14	US-10-156-761-3629	Sequence 3629, Ap
15	66.8	6.9	11058	14	US-10-156-761-3629	Sequence 3629, Ap

16	66.8	6.9	18876	14	US-10-329-079-42	Sequence 42, Appl 1
17	66.8	6.9	61944	14	US-10-329-079-34	Sequence 34, Appl 1
18	64.6	6.6	1230	14	US-10-156-761-5404	Sequence 5404, Ap
19	63.8	6.5	8244	15	US-10-402-842-3	Sequence 3, Appl 1
20	63.8	6.5	47988	15	US-10-402-842-1	Sequence 1, Appl 1
21	62.4	6.4	32768	9	US-09-070-927A-17	Sequence 17, Appl 1
22	62.2	6.4	1227	15	US-10-458-201-23	Sequence 23, Appl 1
23	62.2	6.4	15120	15	US-10-458-201-1	Sequence 1, Appl 1
C 24	61.4	6.3	1230025	15	US-10-289-762-1	Sequence 1, Appl 1
C 25	61.2	6.3	975	10	US-09-938-901-1	Sequence 1, Appl 1
26	60.4	6.2	566	12	US-10-424-599-105239	Sequence 105239,
27	59.8	6.1	1029	10	US-09-938-901-5	Sequence 5, Appl 1
28	59.8	6.1	5355	14	US-10-205-032-19	Sequence 19, Appl 1
29	59.8	6.1	60196	14	US-10-205-032-1	Sequence 1, Appl 1
30	59.4	6.1	272	11	US-09-864-408A-9041	Sequence 9041, Ap
31	58.8	6.0	1122	14	US-10-156-761-4667	Sequence 4667, Ap
32	58.8	6.0	1122	12	US-10-671-403-106	Sequence 106, App
33	58.8	6.0	1128	12	US-10-671-419-106	Sequence 106, App
34	58.8	6.0	1365	14	US-10-156-761-6140	Sequence 6100, Ap
35	57.6	5.9	4437	14	US-10-156-761-4189	Sequence 4189, Ap
C 36	57.4	5.9	1404	14	US-10-156-761-2469	Sequence 2469, Ap
37	57.4	5.9	1590	12	US-10-671-403-3	Sequence 3, Appl 1
38	57.4	5.9	1590	12	US-10-671-419-3	Sequence 3, Appl 1
39	57.4	5.9	2007	12	US-10-671-403-3	Sequence 1, Appl 1
40	57.4	5.9	2007	12	US-10-671-419-1	Sequence 1, Appl 1
41	57.4	5.9	2934	10	US-09-938-901-7	Sequence 7, Appl 1
42	57.2	5.9	1728	14	US-10-156-761-3783	Sequence 3783, Ap
43	57.2	5.9	339630	15	US-10-398-221-7	Sequence 7, Appl 1
C 44	57.2	5.9	3011208	15	US-10-398-221-7	Sequence 2058, Ap
45	57	5.8	11238	14	US-10-305-032-15	Sequence 15, Appl 1

## ALIGNMENTS

```

RESULT 1
US-09-938-901-1
? Sequence 1, Application US/09938901
? Publication No. US2003008291A1
? GENERAL INFORMATION:
? APPLICANT: Kuramitsu Seiki,
? APPLICANT: Yokoyama Shigeyuki
? TITLE OF INVENTION: GENE ENCODING DNA REPAIR ENZYME
? FILE REFERENCE: PH-1261-US
? CURRENT APPLICATION NUMBER: US/09/938, 901
? CURRENT FILING DATE: 2001-08-24
? PRIOR APPLICATION NUMBER: JP2001-47762
? PRIOR FILING DATE: 2001-02-23
? NUMBER OF SEQ ID NOS: 17
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 1
? LENGTH: 975
? TYPE: DNA
? ORGANISM: Thermus thermophilus
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1)..(975)
? US-09-938-901-1

```

```
Db 121 CGGGTGAAGACGCGCTTCCCTATTACCGCGCTTTCTGAGCGCTTCCCACTGAAG 180
Qy 181 GCCCTGGCCGCGCTTCCCTGAGAGAGTCTTAGAGTCTGGAGGGGGCGGCTACTAC 240
Db 181 GCCCTGGCCGCGCTTCCCTGAGAGAGTCTTAGAGTCTGGAGGGGGCGGCTACTAC 240
Qy 241 CGGGGGGGGAAACCTTCCACCGCTGCGCCGGAAGCGTGAAGAGCTTCCCGGAGCTTC 300
Db 241 CGGGGGGGGAAACCTTCCACCGCTGCGCCGGAAGCGTGAAGAGCTTCCCGGAGCTTC 300
Qy 301 GCCGAGCTTGGGGGGCTTCTTGTCTCGAGCTTACACCGCGCGCGGTGGCTTCATC 360
Db 301 GCCGAGCTTGGGGGGCTTCTTGTCTCGAGCTTACACCGCGCGCGGTGGCTTCATC 360
Qy 361 GCCTTGGGGGAGCGGGTGGCGGCGGTGAACGAGAACTTCGAGAGGCTCTCCCGCTTC 420
Db 361 GCCTTGGGGGAGCGGGTGGCGGCGGTGAACGAGAACTTCGAGAGGCTCTCCCGCTTC 420
Qy 421 TTGCGCCGGGAAAGCCCAAGAGAGAGCTTTTCCGCTTGGCCAGGCGCTCTCCCTC 480
Db 421 TTGCGCCGGGAAAGCCCAAGAGAGAGCTTTTCCGCTTGGCCAGGCGCTCTCCCTC 480
Qy 481 GAGGGGTGAGACCCCGGGGGTGTGAACCAAGCCCTCATGAGCTCGGGCCACGGTCTGC 540
Db 481 GAGGGGTGAGACCCCGGGGGTGTGAACCAAGCCCTCATGAGCTCGGGCCACGGTCTGC 540
Qy 541 CTGCGGAAACGCGCCCGTGTGCGGGGCTGCCCTTGAAGGGCTTCTGCGGGGAGAGAG 600
Db 541 CTGCGGAAACGCGCCCGTGTGCGGGGCTGCCCTTGAAGGGCTTCTGCGGGGAGAGAG 600
Qy 601 GCGCCCGGGCGCTACCCCGCGCCAGAGAGCGCGGGGAGAGAGAGAGAGAGAGAGAG 660
Db 601 GCGCCCGGGCGCTACCCCGCGCCAGAGAGCGCGGGGAGAGAGAGAGAGAGAGAGAG 660
Qy 661 CTGCTCTCTCTCGGGCGGAGGGGGTGAACCTGAGAAAGCTTGAAGGGCGCTTCAAGGC 720
Db 661 CTGCTCTCTCTCGGGCGGAGGGGGTGAACCTGAGAAAGCTTGAAGGGCGCTTCAAGGC 720
Qy 721 CTGTAAGCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 780
Db 721 CTGTAAGCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 780
Qy 781 GTGAGGTCTAGAGCCCTTGAAGAGAGTGCACGCGCTTCAACCGAGAGCTTCCGCTG 840
Db 781 GTGAGGTCTAGAGCCCTTGAAGAGAGTGCACGCGCTTCAACCGAGAGCTTCCGCTG 840
Qy 841 GAGGTGCGGGGGGCTTTTGGAGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db 841 GAGGTGCGGGGGGCTTTTGGAGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Qy 901 AAGCTATGAGAGAGTGTCTCCGCAAGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Db 901 AAGCTATGAGAGAGTGTCTCCGCAAGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Qy 961 CCCCTCCCGGACCA 975
Db 961 CCCCTCCCGGACCA 975
```

## RESULT 2

```
US-10-156-761-4691
; Sequence 4691, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
```

```
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; SEQ ID NO 4691
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(939)
US-10-156-761-4691

Query Match 11.5%; Score 112.4; DB 14; Length 939;
Best Local Similarity 52.1%; Pred. No. 3e-15; Mismatches 306; Indels 33; Gaps 4;
Matches 369; Conservative 0; Mismatches 306; Indels 33; Gaps 4;
```

```
Qy 4 GAGGCTGCGGAAAGCCCTCTGCTGTGATCGGAGAAAGCCCGCCCTTCCCTGCG 63
Db 73 GAGGCTGCGAGCCCGAGTATGATGATTCAGACAGACAGCCGCTGACCTCCGTGG 132
Qy 64 CGGG---GAGAAAGACCTTACCGGCTCTGCTCCGAGTCTTTCAGACAGACC 120
Db 133 CGCGCCCGGAGCGCGGCTCTGAGGCTATGATGATTCAGACAGACAGCCGCTGAC 192
Qy 121 CGGGTGAAGAGCGCTTCCCTTATTCAGCGCGCTTCTGAGCGCTTTCACCTGAAG 180
Db 193 CGGGTGAAGAGCGCGCTTCCCTTATTCAGAGAGTGTGAGAGAGAGAGAGAGAGAG 252
Qy 181 GCGCTGCGCGGCTTCCCTGAGAGAGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 253 GACCTCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 312
Qy 241 CGGGGGGGAACCTTCAACCGCTTGGCC-----GAGAGCTGAGAGAG 285
Db 313 CGGGCGCGCTGCGCTGCGCAAGCGCGCGCTGCGCATTAAGAGAGAGAGAGAGAG 372
Qy 313 CTTCCCGGAGCTTCCGAGCTTCCGGGGCTTCTGCTGTGCTTCCGCTTCAACCGCGCG 345
Db 373 GTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 432
Qy 346 GCGGTGCTTCATGCTCTTCCGAGAGCGGAGTGGCGCGGTGAGAGAGAGAGAGAGAG 405
Db 433 GCGGTGCTGCTTCCCTTCAAGCGCGCAAGCGCGCTTCTGAGACCAAGTCCGCGGT 492
Qy 406 GTCCTTCCCGGCTTTCGCTCCGAGAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAG 459
Db 493 GTCCTTCCCGGCTTTCGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 552
Qy 460 -----CTGCGCGAGGCTTCTCCCGAGGCGTGAACCGGGG---TGTGAACAG 510
Db 553 CGGAGCTGCGCGCGGCTGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 612
Qy 511 GCGCTATGAGAGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570
Db 613 GCGCTATGAGAGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672
Qy 571 CCCCTAAGGGCTTCTGCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 630
Db 673 CGGATGCGCGGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 732
Qy 631 CGCGGGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db 733 CGCGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
```

## RESULT 3

```
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
```



OY 579 GGCCCTTCGCCGG 592  
DB 1049 GAGCCTGTGCCGG 1062

## RESULT 5

US-09-974-300-1583  
; Sequence 1583, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 10085-500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq For Windows Version 4.0  
; SEQ ID NO 1583  
; LENGTH: 528  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(528)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-974-300-1583

## Query Match

Best Local Similarity 9.5%; Score 92.8; DB 9; Length 528;  
Matches 227; Conservative 0; Mismatches 183; Indels 15; Gaps 1;

OY 17 AAGCCCTCTCGCTGTACCGGAAACGCGCCCTCCCTCGGGGGAGAAAG 76  
DB 43 AAGATTTATTTCTGTGTATGACAAAGAAAGGGATCGCTTGCGTACAGCG 102  
OY 77 ACCCTTACCGCGCTCGTGTCTCGAGGTCTTTCGACAGACCCGGGTGAGACAGCC 136  
DB 103 ACCCTTATAGGT 162  
OY 137 TCCCTATTACCGCGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 196  
DB 163 TTTCTTACTTCAACAATTATGAAAAGTTTTCGACCGTGTGAGCGCTGTGAAAGCG 222  
OY 197 CCTGTGAAGAGTCTTGTAGGGTCTGACAGGGCGGGCTTACTACCGCGGGCGAAAC 256  
DB 223 ACGAAGAAAAGTCTTAAAGCTGTGGAAGACTGGGATGCTTCAAGGGTTCCGAAG 282  
OY 257 TCCAGCGC-----CTGGCCCGAAGCGTGTGAGAGCTTCCCGAGCTTCG 301  
DB 283 TGCAGAGCGCTGTACAGGAAGTTACGAACGATCGAGTGTGTCTCTTCAAAAG 342  
OY 302 CCGAGCTTGTGGGGGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 361  
DB 343 AAGAAATTTGGGCACTTAAAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 402  
OY 362 CCTTGGGAGAGCGGT 421  
DB 403 CCTTACATAGCGCGGT 462  
OY 422 TCGCC 426  
DB 463 TTTCC 467

RESULT 6  
US-10-425-114-32132  
; Sequence 32132, Application US/10425114

; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 32132  
; LENGTH: 1660  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMFLB73265D07\_FU1  
US-10-425-114-32132

Query Match 9.1%; Score 89.2; DB 12; Length 1660;  
Best Local Similarity 50.7%; Pred. No. 2.4e-10;  
Matches 282; Conservative 0; Mismatches 233; Indels 21; Gaps 2;

OY 60 CTGCGGGGGAGAGAACCTTACCGCGTCTGTCTCGAGGTCTTCTGACAGAC 119  
DB 278 CGGACGAGAGAGAGAGCGGTACGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 337  
OY 120 CCGGTGAGAGAGCGCTTCCCTTATACCGCGCTTCTGTGAGGCTTTCCACCTGAA 179  
DB 338 GGGGTGCGCGT 397  
OY 180 GGCCTGTGCGCGCTTCTGTGAGAGGTCTTGTAGGTCTGTGAGAGGCGGCTACTA 239  
DB 398 AAGCTGTGCGCGCTTGTGAGAGGTGTGAGAGGTGTGAGAGGTGTGAGAGGTGTG 457  
OY 240 CCGCGGGCGGAACCTTC-----ACCGCTGCGCGGAAGCGTGAAGAGCTTCC 290  
DB 458 CCGAGGCTGTGATTTCTGTGTGAGAGGAGCAAGCAATCATGAAAGGGTTGTTCC 517  
OY 291 CCCGAGCTTGTGCGAGGTGTGCGGGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 350  
DB 518 TTGACAGCGCTTGTGAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 577  
OY 351 GGCCTCATGCTTGTGAGAGCGGTGTGAGAGGTGTGAGAGGTGTGAGAGGTGTG 410  
DB 578 CGTTCAATTGTCTTCAACGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 637  
OY 411 CTCCGCTCTGTGCGCGCGGAAAGCCCAAGAGAA-----GAGCTTTTGGC 458  
DB 638 CAGAGGCTTTCACCAATTGTGTGACAAACCAAGGAATCTCAACAGTGAAGATTCTG 697  
OY 459 CTTGCGCCAGGCTCTCTCCCGAGGCGTGTGACCGGGGTGTGAGACAGGCGCTCAT 518  
DB 698 GGAAGCTGT 757  
OY 519 GGAAGCTGTGAGCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 578  
DB 758 GGAAGTGTGAGCAAACTTATGTAGCAAAAGCTGTGTGTGTGTGTGTGTGTGTGT 817  
OY 579 GGCCTTGTGCGGGGG 594  
DB 818 TAGCTACTGCCAAGCG 833

RESULT 7  
US-10-216-817-31  
; Sequence 31, Application US/10216817  
; Publication No. US20030129619A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE

```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; TITLE OF INVENTION: IN GENES OF THE MULT FAMILY
; FILE REFERENCE: 03495,0233-00000
; CURRENT APPLICATION NUMBER: US/10/216,817
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 912
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-216-817-31

```

```

Query Match      8.9%; Score 86.6; DB 14; Length 912;
Best Local Similarity 50.2%; Pred. No. 9.6e-10;
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

```

```

Qy 22 CTCCTCGCTGTGATCCGGAAACGCGCCCTCCCTGCGGGG---GAGAAGGAC 78
Db 61 CTTCTCGCTGTGATCAGCATCGACCGGATCGCTGCGAGACCGCGGCTGTACG 120
Qy 79 CTTACCGCGTCTGTGATCCGAGGCTCTTTCGACAGACCCGGGTGAGACGCCCTC 138
Db 121 CCGTGGCAGATCTGTGATCAGCATCGATTCGTCGACGACGCGCGCGGGTGTG 180
Qy 139 CCTTATTAACCGCGCTTCTGTGAGAGCTTTCCACCTGAAAGCCCTGCGCGCTTC 198
Db 181 GCGATTCGCGGACGTGGGTGCGGGGTGCGGCGGCGGCGGCGGCGGCGGCGG 240
Qy 199 CTGAAGAGTCTTAAAGGTCTGCGAGGGGCGGCTCTACACCGCGGGCGGAACCTC 258
Db 241 ACCGCGATGTGTTACGCGCTGGGCGAAGCTGGCTATCCAGCGAGCCAGGCTTA 300
Qy 259 CACGCGCTGGCGGAGCGTG-----GAGAGCTTCCCGAGCTTGCC 303
Db 301 CACGATGCGCGACCGTATCGCCCGCGACCAATGACGTGTGCCGACGATATCGAG 360
Qy 304 GAGCTTGGGGGCTTCTGTGATCTGGGCTTACACCGCGGCGGCGGCTTCATCGCC 363
Db 361 ATCTGTGTACCTTGTGCGGCGGTGCGGAGCTACACCGCGCGCGGTGCTTCT 420
Qy 364 TTGGGGAGCGGGGTGCGGCGGTGAGACGGAACGTCCGGAAGGTCTCTCCGCTTT- 422
Db 421 TACGCGCAGCGGGGTGCGGTGTGACACCAATGTGCGGCGGTGTGCGCGCTT 480
Qy 423 -----CGCCCGGAAAGCCCAAGAGAGAGAGCTTTTCCCTCGCCAGGAC 471
Db 481 CAGGCGCGCGCGACGCGCGGTGCGCATCGGTGCGGCGGACACAGCGGATGTGCG 540
Qy 472 CTCCTCCCGAGGGCGTGTGACCCGGGGGTGTGAACACAGGCTCATGAGCTCGGGCC 531
Db 541 CTGTGTGCGCACCGCGAGACGCGCTGAATTTTCGTGCGCGCTGATGAGTGTGGTCG 600
Qy 532 ACGGTTCGCTGTGCGAAGACGCGCGGTGCGGGGCTGCGCCCTAGAGGCGCTTTCGCG 591
Db 601 ACGGTGTGACCGCGCGGACACCGCGGTGCGGGTATATCCGCTGTGACTGTGCGATGG 660
Qy 592 GGAAGAGAGCGCGCGGCGCTACCCCGCGCGCAAGAGCGCGGCGAAGAGAGCGC 651
Db 661 CGGATGCGGGTATTCGCGCGGTGCGAGGTCC---GCCGCGCGGGGCGAGGCTTACAC 717
Qy 652 CTCGTGCGCTGTGCTCTCTCTGCGGCGG 678
Db 718 GGAACCGACCGCAAGTCCGCGAGCG 744

```

RESULT 8  
US-10-216-817-21

```

; Sequence 21, Application US/10216817
; Publication No. US20030129619A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; TITLE OF INVENTION: IN GENES OF THE MULT FAMILY
; FILE REFERENCE: 03495,0233-00000
; CURRENT APPLICATION NUMBER: US/10/216,817
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1312
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-216-817-21

```

```

Query Match      8.9%; Score 86.6; DB 14; Length 1312;
Best Local Similarity 50.2%; Pred. No. 8.9e-10;
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

```

```

Qy 22 CTCCTCGCTGTGATCCGGAAACGCGCCCTCCCTGCGGGG---GAGAAGGAC 78
Db 261 CTTCTCGCTGTGATCAGCATCGACCGGATCGCTGCGAGACCGCGGCTGTACG 320
Qy 79 CTTACCGCGTCTGTGATCCGAGGCTCTTTCGACAGACCCGGGTGAGACGCCCTC 138
Db 321 CCGTGGCAGATCTGTGATCAGCATCGATTCGTCGACGACGCGCGGGTGTG 380
Qy 139 CCTTATTAACCGCGCTTCTGTGAGAGCTTTCCACCTGAAAGCCCTGCGCGCTTC 198
Db 381 GCGATTCGCGGACGTGGGTGCGGGGTGCGGCGGCGGCGGCGGCGGCGGCGG 440
Qy 199 CTGAAGAGTCTTAAAGGTCTGCGAGGGGCGGCTCTACACCGCGGGCGGAACCTC 258
Db 441 ACCGCGATGTGTTACGCGCTGGGCGAAGCTGGCTATCCAGGCGAGCGAGGCTTA 500
Qy 259 CACGCGCTGGCGGAGCGTG-----GAGAGCTTCCCGAGCTTGCC 303
Db 501 CACGATGCGCGACCGTATCGCCCGCGACCAATGACGTGTGCCGACGATATCGAG 560
Qy 304 GAGCTTGGGGGCTTCTGTGATCTGGGCTTACACCGCGGCGGCGGCTTCATCGCC 363
Db 561 ATCTGTGTACCTTGTGCGGCGGTGCGGAGCTACACCGCGCGCGGTGCTTCT 620
Qy 364 TTGGGGAGCGGGGTGCGGCGGTGAGACGGAACGTCCGGAAGGTCTCTCCGCTTT- 422
Db 621 TACGCGCAGCGGGGTGCGGTGTGACACCAATGTGCGGCGGTGTGCGCGCTT 680
Qy 423 -----CGCCCGGAAAGCCCAAGAGAGAGCTTTTCCCTCGCCAGGAC 471
Db 681 CAGGCGCGCGCGACGCGCGGTGCGCATCGGTGCGGCGGACACAGCGGATGTGCG 740
Qy 472 CTCCTCCCGAGGGCGTGTGACCCGGGGGTGTGAACACAGGCTCATGAGCTCGGGCC 531
Db 741 CTGTGTGCGCACCGCGAGACGCGCGCTGAATTTTCGTGCGCGCTGATGAGTGTGGTCG 800
Qy 532 ACGGTTCGCTGTGCGAAGACGCGCGGTGCGGGGCTGCGCCCTAGAGGCGCTTTCGCG 591
Db 801 ACGGTGTGACCGCGCGGACACCGCGGTGCGGGTATATCCGCTGTGACTGTGCGATGG 860
Qy 592 GGAAGAGAGCGCGCGGCGCTACCCCGCGCGCAAGAGCGCGGCGAAGAGAGCGC 651
Db 861 CGGATGCGGGTATTCGCGCGGTGCGAGGTCC---GCCGCGCGGGGCGAGGCTTACAC 917
Qy 652 CTCGTGCGCTGTGCTCTCTCTGCGGCGG 678
Db 918 GGAACCGACCGCAAGTCCGCGAGCG 944

```

```
RESULT 9
US-10-158-844-12/c
; Sequence 12, Application US/10158844
; Publication No. US20040029118A1
; GENERAL INFORMATION:
; APPLICANT: Kunach et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude Pentium 3
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/158,844
; FILING DATE: 03-Jun-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/961,527
; FILING DATE: 1997-10-30
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-158-844-12

Query Match      8.4%; Score 82; DB 12; Length 9909;
Best Local Similarity 52.2%; Pred. No. 5.7e-09;
Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

QY 22 CTCCTGCTGTAACCGGAAACGCGCCCTCCCTGGCGGGGAGAAAGACCT 81
DB 5808 CTCTCCCTGGTATGATGAAACAAAGATTGCTTGGAGAGAAAGTAAATCCT 5749
QY 82 TACCGGCTGCTGTCCTCCAGGTCCTTTCGACAGACCCGGGTGAGACGCCCTCCC 141
DB 5748 TATCAATCTGGGTATCTGAATTCATGCTTTCAGACAGACCGGGTGTATACGTTATCCCT 5689
QY 142 TATTACCGCGCTTTCGTGAGCGCTTCCACCTGAAAGCCCTGGCGCGGCTTCCCTG 201
DB 5688 TACTACGAAAGATTTCGACTGGTTTCCACTGTTCGAAAGTCTGGCACTGGCCCTGAG 5629
QY 202 GAAGAGCTCTTAGGCTTCGACAGGGGCGGCTA-----CTACCGCGG 246
DB 5628 GAGAGTTTCTGAAGACTTGGAGGGGCTTGGGCTATATTCTCGAGTTCCGAATATGACAG 5569
QY 247 GCGAAACACTTCACCGCGCTGCGCGGAGCGTGGAGAGCTTCCCGGAGCTTGGCGAG 306
DB 5568 GCTGAGCGCCAGCAGATATGACTGACTTGTGTGGCCAAATTTCCAAATACCTTATGAGGA 5509
QY 307 CTTGCGGGGCTTCCGTGCTCGGGCTTACACCGCGCGCGGTGCGCTTCATGCGCTTC 366
DB 5508 ATTTCAGCTTGAAGAGGATTTGACCTTACACAGAGAGCCATTTCCAGTATTTGCTTTT 5449
QY 367 GGGGAGCGGTGGCGCGGTGAGACGGGAAGTCCGAGAGGCTCTTCGCCCTTTTCG 424
```

```
DB 5448 AACTTGCTGAGCCACGCTGATGTGTAATGATCATGCGGGCTTTGGCGCGTCTGTTG 5391

RESULT 10
US-09-864-866-46
; Sequence 46, Application US/09864866
; Patent No. US20020127656A1
; GENERAL INFORMATION:
; APPLICANT: Lloyd, R. Stephen
; APPLICANT: Nguyen, Khao K.
; TITLE OF INVENTION: DNA REPAIR POLYPEPTIDES AND METHODS OF USE
; FILE REFERENCE: 265,001,701,01
; CURRENT APPLICATION NUMBER: US/09/864,866
; PRIOR FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 46
; LENGTH: 968
; TYPE: DNA
; ORGANISM: Micrococcus luteus
US-09-864-866-46

Query Match      8.3%; Score 81.2; DB 9; Length 968;
Best Local Similarity 47.2%; Pred. No. 1.3e-08;
Matches 300; Conservative 0; Mismatches 318; Indels 18; Gaps 1;

QY 44 ACGCCGCCCTCCCTCCCTGCGGGGAGAGACCTTACCGGCTCTGCTTCGAG 103
DB 200 ACCGTAACGCGTGGCGGAGCTGACTTGAACGCCCTTTCAGCTGCTGTGGCAGCG 259
QY 104 TCCCTTCGACGACACCGCGGTGAGACAGCCCTCCCTATTATCCCGCTTTCGAGC 163
DB 260 TCGTTCGCCCGCCAGACCAACGACGTGCGGTCAACGACGACGCGCGGTTCGCC 319
QY 164 GCTTTCACACCTGGAAGCCCTGCGCGCGCTTCCCTGAGAGAGTCTTATAGGTTCG 223
DB 320 GCTTCCCGGATGCCACGCGGATGCGCGGCGCACCGACCGGAGCTGATGTC 379
QY 224 AGGGGGGGGCTATACCGG-----CGGCGGAAACCTCCACCGCC 265
DB 380 GCTTCACGGGGTTCATCCGGAACAGGCTCCGAGATCTCGGCTGCCAGAGCTTCG 439
QY 266 TGGCCCGAAGCGTGAAGAGCTTCCCGGAGCTTGGCGAGCTTCCGGGGCTTCGTG 325
DB 440 TGGGCGGACAGAGCGGAGGTCCTCCGCTTCGAGAGACTTCGTGGCGCTCCGGGG 499
QY 326 TGGGCTTACACCGCGCGCGGTGAGCTTCATGCGCTTTCGAGAGCGGGGTGGCGCG 385
DB 500 TGGGCGGAAAGCGCGCTTTCGTGTGCTCGGCAACGCTTCGCGCAGCCGGGATACCG 559
QY 386 TGAACGGGAAGTTCGAGAGGCTTCTCCCGGCTTTCGCGGAAAGCCCAAGAGA 445
DB 560 TGAACGACACTTGGCGCGGTCCTCCGCGCTTGGGTTTCAGAGAGAGACCGACCG 619
QY 446 AGAGCTTTTGGCTTCCCGGAGGCTTCTCCCGAGGCGGTGACCGCGGGGTGTGA 505
DB 620 GTAAAGTTCAGACCGCGGTGGCGGCTTTCCTCCCGCGGAGCTGGAAGATCTCT 679
QY 506 ACCAGGCTTCATGAGACTTCGAGGCGCAGAGTCTGCTGCGAAAGGCGCCCTTGGAG 565
DB 680 CCGACCGGCTGATCTTCAGAGCGCGCGGTGTGCCACGCGCGCGCGCGGTGCGGGC 739
QY 566 CTTGCGCCCTTGAAGGCGCTTTCGCGGGGAAAGAGCCCGCGGCTTACCGCGCGCA 625
DB 740 GGTGCTCGAATGCGCGCTGTGTGCTCTTACGCGCGGGGAGACGACCCCGAGCGGG 799
QY 626 GGAAGCGCGCGGAGAGAGAGCGCTGTGCGCC 661
DB 800 CGGCGCGCTGCTGCTTCAAGCTTAAAGCCCGGC 835
```

RESULT 11  
US-10-156-761-4577  
; Sequence 4577, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 4577  
; LENGTH: 930  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(930)  
US-10-156-761-4577

Query Match 7.3%; Score 71.4; DB 14; Length 930;  
Best Local Similarity 47.1%; Pred. No. 1.7e-06;  
Matches 305; Conservative 0; Mismatches 321; Indels 21; Gaps 2;  
QY 39 GAAAAAGCCCGCCCTCCCTGCGGGGGGAGAGACCTTAACCGCGTCTGCTC 98  
DB 207 GGTGTACCCGTAACCAACCCGAGTTGACTTGAGAACTCTTCCAGCTGATCTTGGC 266  
QY 99 CGAGGTCTTCTGAGAGACCGGGGTGAGCAGCCCTCCCTTAATTAACCGCGCTTCT 158  
DB 267 GACCGTCTGTGCGCGAGACCAACCACTGCGGGTGAACCAAGACGCGCGCTGTT 326  
QY 159 GGAAGCGCTTCCACCCCTGAAGCGCTGCGCGGCTTCTTCTGAAGAGTCTTAAGG 218  
DB 327 CGCNAATACCGACCCCGAGACCTGCGCGGCCCAACCGAGAGTCTGAGAGAT 386  
QY 219 CTGGCAGGGGGGGGCTAATACCGCGG-----GGCGAACAATTCCA 260  
DB 387 CTTCCGCGCCCAACCGGCTTCTCCGGGCCAAGACCAAGTCGTATAGGGCTGTCCAAG 446  
QY 261 CCGGCTGCGCGAGCGGTGAGAGCTTCCCGAGCTTCCCGAGCTTCCGGGGCTTCC 320  
DB 447 CTTGTGAGAGATTGCGCGGCGAGGTCCCGCGGCTTCAAGATCTCTCAAGCTGCC 506  
QY 321 TGTCTCGGCTTCAACCGCGCGCGGTGAGCTTCAATCGCTTCCGGGAGCGGGTGC 380  
DB 507 CGGCTGAGACGAAAGCGGCTTCTGTGTCTGCGAAGCGCTTCCGGCGCGCGCAT 566  
QY 381 GCGGCTGAGCGGAAAGTCCGAGAGGTCTTCTCCGCTCTTCCCGGGAAGCCCCAA 440  
DB 567 CACCGTGAGACCACTATTTCCAGCGGCTGTCCGCGCTGAGAGTGAACCGAGAAAGA 626  
QY 441 GGAAGAGAGCTTTTCCGCTTCCGCGCGCGGCTCTCCCGAGGGCGTGAACCGGGGGT 500  
DB 627 TCCGACAAAGATCGAGGGCGGCTGTCCGCGCTTCTCCGAGAGCGATGAGAGATCT 686  
QY 501 GTGAAACAGGCGCTTCAATGAGTCTGCGGCGCAACGCTGCTGCGGAAACGCGCGTGG 560  
DB 687 CTGCGACCATG---TGATCTTCAACGCGCGCGCATGTGTACGCGCGCAAGCGGGCTG 743  
QY 561 CGGGGCTGCGCCCTAGGGGCTTCTGCGCGGGGAGAGAGGCGCCCGGGCGCTACCCGCG 620

DB 744 CGGGGCTGCGCCCATTCGCGCGCGCTTCTGCGCGCTACGGGAGGCGAGACCGCGA 803  
QY 621 GCCCAGAGAGCGCGGCGAGAGAGCGGCTGTGCGCGCTGCTC 667  
DB 804 GAAGCGAAGAGCTCTGAAGTACGAGAGGCGGCTTCCCGCGCC 850

RESULT 12  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 7.3%; Score 71.4; DB 14; Length 9025608;  
Best Local Similarity 47.1%; Pred. No. 2.6e-07;  
Matches 305; Conservative 0; Mismatches 321; Indels 21; Gaps 2;  
QY 39 GAAAAAGCCCGCCCTCCCTGCGGGGGGAGAGACCTTAACCGCGTCTGCTC 98  
DB 5607423 GGTGTACCCGTAACCAACCCGAGTTGACTTGAGAACTCTTCCAGCTGATCTTGGC 5607482  
QY 99 CGAGGTCTTCTGAGAGACCGGGGTGAGCAGCCCTCCCTTAATTAACCGCGCTTCT 158  
DB 5607483 GACCGTCTGTGCGCGAGACCAACCACTGCGGGTGAACCAAGACGCGCGCTGTT 5607542  
QY 159 GGAAGCGCTTCCACCCCTGAAGCGCTGCGCGGCTTCTTCTGAAGAGTCTTAAGG 218  
DB 5607543 CGCNAATACCGACCCCGAGACCTGCGCGGCCCAACCGAGAGTCTGAGAGAT 5607602  
QY 219 CTGGCAGGGGGGGGCTAATACCGCGG-----GGCGAACAATTCCA 260  
DB 5607603 CTTCCGCGCCCAACCGGCTTCTCCGGGCCAAGACCAAGTCGTATAGGGCTGTCCAAG 5607662  
QY 261 CCGGCTGCGCGAGCGGTGAGAGCTTCCCGAGCTTCCCGAGCTTCCGGGGCTTCC 320  
DB 5607663 CTTGTGAGAGATTGCGCGGCGAGGTCCCGGGCGCTTCAAGATCTCTCAAGCTGCC 5607722  
QY 321 TGTCTCGGCTTCAACCGCGCGCGGTGAGCTTCAATCGCTTCCGGGAGCGGGTGC 380  
DB 5607723 CGGCTGAGACGAAAGCGGCTTCTGTGTCTGCGAAGCGCTTCCGGCGCGCGCAT 5607782  
QY 381 GCGGCTGAGCGGAAAGTCCGAGAGGTCTTCTCCGCTCTTCCCGGGAAGCCCCAA 440  
DB 5607783 CACCGTGAGACCACTATTTCCAGCGGCTGTCCGCGCTGAGAGTGAACCGAGAAAGA 5607842  
QY 441 GGAAGAGAGCTTTTCCGCTTCCGCGCGCGGCTCTCCCGAGGGCGTGAACCGGGGGT 500





APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 3629  
LENGTH: 11058  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(11058)  
US-10-156-761-3629

Query March 6.9%; Score 66.8; DB 14; Length 11058;  
Best Local Similarity 48.9%; Pred. No. 9.6e-06;  
Matches 278; Conservative 0; Mismatches 272; Indels 18; Gaps 3;  
QY 331 CTTTACACCGCGCGCGCGCTGCTTCATCGCTTTGCGGAGACCGGCGCGGTGAC 390  
DB 6622 CCGTCGCGGTGCAAGATCTACCGCGCTTCAGAGCGGGCGCGCGCGCTCGAG 6681  
QY 391 GGGAACTCCGAGAGGTCTCTCCGCTTTGCGCGGAAAGCCCAAGAGAGAG 450  
DB 6682 GAGTTCACTGCGCCGCGGAACGCGCGCACTGTTCACTGAGAGCGGCGCGCTGCGG 6741  
QY 451 CTTTTCGCGCTGCGCGAGGCGCTCTCCCGAGGCGTGAACCGGGGGTGTGAACAG 510  
DB 6742 ATGACCGCACAGTGAAGAGAGAGAGCGCGCTCGCGCTTCAGAGTGTG--CCAC 6798  
QY 511 GCCCTCATGAGCTCGGGGCGACAGGTCTGCTGCGGAACGCGCGCGTTCGCGGGCTGC 570  
DB 6799 GCCATCACGAGAGGCTGAGCAACCGCGCGCTGCTGATGAGAGCTGTCAGCAATACCGG 6858  
QY 571 CCCCTAGGGGCTTCTGCGGGGAGAGAGCCCTCGGCGCTACCCGCGCGAGAG 630  
DB 6859 CGGCTGCGCGACGCGGCGACAGCCGTCGAGAGCCCGCGCGCGCTACCGGAG 6918  
QY 631 CGCGGCGAGAGAGAGCGCTGCTGCGCTCTGCTGCGGCGGAGAGGGGTGCAC 690  
DB 6919 TTCTGTCGACAGAGTGCCTCTGAGTGGAGCAGACCGGCGTACTGGGCTGAC 6978  
QY 691 CTGAAAGGCTTGAAGGGCGCTTCCAG-----GGCTCTACGCGCTGCCCTTTTC 742  
DB 6979 GTCAACCGCGCGCTTTCGCGCGCTTTCGCGCTGCGCGCGCGCGCGAGCGCTCGGC 7038  
QY 743 CCCCTGAGAGCTTCCCGGCGGAGAGCG-----CCTTGGGGTGAAGTCTAGGCC 795  
DB 7039 CCGCGGAGACTTTCGCGGTGCGGTGAGCTGCGCCCTTTCAGACAGGGGTGCGCGCC 7098  
QY 796 CTAGCGGAGGTGCGCAAGCGCTCAACGAGAGGCTTTCGCTGAGAGGTGCGGGGGCC 855  
DB 7099 CTGACCGCGCGCGCGCGGTCTCTCTCAAGAGCTGTGCTGCGCGCGCACTGAAGCG 7158  
QY 856 CTTTGGAGAGGAGGAGGAGAGCCCT 883  
DB 7159 CTGAGCGTGTGACCGAGAGAGACCGCT 7186

Search completed: March 4, 2004, 07:02:34  
Job time : 1286 secs

**This Page Blank (uspto)**



QY 400 CGAGGGCTCTCTCCCGCTTTGCGCCGGGAACCCCAAGAGAGAGCTTTGCCC 459  
Db 413 AAGCGGTACTGCGCGGCTACTGCGGAGAGAGCTATCCCGGGAACCAAGGTGCCC 472  
QY 460 CTGCGCCAGAGGCTCTCTCCCGAGGCGTGAACCCGGGGGTGTGG-----AAC 507  
Db 473 AGGGCGCTGTGGAGAGCCGCGAAGCTTCAACCCCGACGACGAGGTATACCATACAC 532  
QY 508 CAGGCGCTCATATGAGCTGAGGGGCAAGGTCTGCTGCGGAAACGAGCCCGTTGGGGGCC 567  
Db 533 CAGGGGATGATGAGACTTGGGCGCACCTCTGACGCGCAGCAAGACCAAGTTGCTGCTT 592  
QY 568 TGCCCGCTAGGGGCTTCTGCGGGGGGAA 596  
Db 593 TGCCCGCTGTCTCTCCGCTGCGCGCGCA 621

RESULT 2  
US-09-252-991A-6886/c  
; Sequence 6886, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6886  
; TYPE: DNA  
; LENGTH: 1401  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6886

Query Match 12.9%; Score 125.4; DB 4; Length 1401;  
Best Local Similarity 54.7%; Pred. No. 2,2e-16;  
Matches 311; Conservative 0; Mismatches 231; Indels 27; Gaps 2;  
QY 55 CTCCCTGCGGGGAGAGAACCTTACCGCGCTGTGCTCCGAGTCTTCTGCAG 114  
Db 1137 CTCCCTGCGGAGAGAGGACATCACTCCCTACCGGCTGTGCTCGGAATCATGCTGCAG 1078  
QY 115 CAGACCGGGTGAAGAGCGCTTCTTATACCGCGCTTTTGGAGCGTTTCCACC 174  
Db 1077 CAGACCGAGTCAAGACCGCTGTGCTTCTTCACTTCACTGACCGTTTGCCTGCAG 1018  
QY 175 CTGAAGCCCTGCGCGGCTTCTCTGAAGAGTCTTAAAGGTCTGCGAGGGGGGCG 234  
Db 1017 GTCGAGGCACTGCGCGGCGGCGGCGGAGAGCAAGTCTCTACCTGTGACCGGCTCGGC 958  
QY 235 TACTACCGGCGGCGGAACCTTCAACCGCTGCGCGCGGAAGCT-----G 279  
Db 957 TACTACAGCGGTGCGCGCAACTGCAAGACCGGCGAGATGTGTGAGCGGATCG 898  
QY 280 GAGGAGCTTCCCGGAGCTTTCGCGCGGCTTCTGCGGCTTCTGCGGCTTACACC 339  
Db 897 GGGGAGTTCCCGCGGAGCTTCAAGCACTGCGCGAATGCTCGGCGATGCGCGCTCCACC 838  
QY 340 GCGGCGGCGGTGAGCTTCACTGCGGAGCGGCGGAGCGGCGGTGAGCGGAGCGTC 399  
Db 837 GCTGAGGCTATGCGCGAGCTGTGATGAGGCTGCGCGGACCGAGTCTTGAAGGCAAGCTC 778  
QY 400 CGGAGGCTCTCTCCCGCTTTCGCGCGGAGAAACCCCAAGAGAGAGCTTTGCCC 459  
Db 777 AAGCGGTACTGCGCGGCTACTGCGGAGAGCGGCTATCCCGGGAACCAAGGTGCCC 718  
QY 460 CTGCGCCAGAGGCTCTCTCCCGAGGCGGTGAGCCCGGGGGTGTGG-----AAC 507

Db 717 AGGGCGCTGTGGAAAGCCCGGAGCTTACCCCGGACGAGCGGCTCAACACTACACC 658  
QY 508 CAGGCGCTCATATGAGCTGAGGGGCAAGGTCTGCTGCGGAAACGCGCGCTTTGGGGGCC 567  
Db 657 CAGGCGATATGAGCTTGGCGCGCACCTCTGACGCTGACAGCGCAGCAAGCCAGTTGCTGCTT 598  
QY 568 TGCCCGCTAGGGGCTTCTGCGGGGGGAA 596  
Db 597 TGCCCGCTGTCTCTCCGCTGCGCGCGCA 569

RESULT 3  
US-09-252-991A-6739  
; Sequence 6739, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6739  
; TYPE: DNA  
; LENGTH: 1413  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6739

Query Match 12.9%; Score 125.4; DB 4; Length 1413;  
Best Local Similarity 54.7%; Pred. No. 2,2e-16;  
Matches 311; Conservative 0; Mismatches 231; Indels 27; Gaps 2;  
QY 55 CTCCCTGCGGGGAGAGAACCTTACCGCGCTGTGCTCCGAGTCTTCTGCAG 114  
Db 409 CTGCGCTGAGAGAGAGGACATCACTCCCTACCGGCTGTGCTCGGAATCATGCTGCAG 468  
QY 115 CAGACCGGGTGAAGAGCGCTTCTTATACCGCGCTTTTGGAGCGTTTCCACC 174  
Db 469 CAGACCGAGTCAAGACCGCTGTGCTTCTTATGACGTTTATGACAGCGTTGCGGAG 528  
QY 175 CTGAAGCCCTGCGCGGCTTCTCTGAAGAGTCTTAAAGGTCTGCGAGGGGGGCG 234  
Db 529 GTCGAGGCACTGCGCGGCGGCGGCGGAGAGCAAGTCTTCACTGTGAGCGGCTCGGC 588  
QY 235 TACTACCGGCGGCGGAACCTTCAACCGCTGCGCGCGGAAGCT-----G 279  
Db 589 TACTACGCGGTGCGCGCAACTGCAAGACCGGCGAGATGTGTGAGCGGATCG 648  
QY 280 GAGGAGCTTCCCGGAGCTTTCGCGCGGCTTCTGCGGCTTCTGCGGCTTACACC 339  
Db 649 GGGGAGTTCCCGCGGAGCTTCAAGCACTGCGCGAATGCTCGGCGATGCGCGCTCCACC 708  
QY 340 GCGGCGGCGGTGAGCTTCACTGCGGAGCGGCGGAGCGGCGGTGAGCGGAGCGTC 399  
Db 709 GCTGAGGCACTGCGCGAGCTGTGATGAGGCTGCGCGGACCGAGTCTTGAAGGCAAGCTC 768  
QY 400 CGGAGGCTCTCTCCCGCTTTCGCGCGGAGAAACCCCAAGAGAGAGCTTTGCCC 459  
Db 769 AAGCGGTACTGCGCGGCTACTGCGGAGAGCGGCTATCCCGGGAACCAAGGTGCCC 828  
QY 460 CTGCGCCAGAGGCTCTCTCCCGAGGCGGTGAGCCCGGGGGTGTGG-----AAC 507  
Db 829 AGGGCGCTGTGGAAAGCGCGGAAGCTTCAACCCCGACGACGAGGTCAACCATACAC 888  
QY 508 CAGGCGCTCATATGAGCTGAGGGGCAAGGTCTGCTGCGGAAACGCGCGCTTTGGGGGCC 567  
Db 889 CAGGCGATATGAGCTTGGCGCGCACCTCTGACGCGCAGCAAGACCAAGTTGCTGCTT 948

Qy 568 TGCCCCCTAGGGGCTTCTGCGGGGGA 596  
 Db 949 TGCCCGCTGCTCTCCGGCTCGCGCGCA 977

RESULT 4

US-09-489-039A-4788  
 ; Sequence 4788, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709,2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; PRIOR FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 4788  
 ; LENGTH: 1551  
 ; TYPE: DNA  
 ; ORGANISM: Klebsiella pneumoniae  
 ; US-09-489-039A-4788

Query Match 11.6%; Score 113; DB 4; Length 1551;

Best Local Similarity 53.0%; Pred. No. 6,1e-14;  
 Matches 307; Conservative 0; Mismatches 245; Indels 27; Gaps 2;

Qy 53 CCCCTCCCTGGCGGGGAGAGACCCTTACCGGCTCTGCTCGAGGCTCTTCCG 112  
 Db 560 CCTGCTCCGCAATGCGCAAGACGCTTCAAAAGTAGGCTCTCCGAAGTAGTTGC 619  
 Qy 113 AGCAGACCCGGGTGAGAGAGCCCTCCCTATTACCGCGCTTCTGAGGCTTTCCCA 172  
 Db 620 AAGAACCCAGGTACACACCGTATCCCTATTGAAAGCTTTATGCGGCTTTCCCA 679  
 Qy 173 CCTGAAAGCCCTGCGCGGCTTCTGAGAGGCTCTTGAAGGCTCTGAGAGGCGCG 232  
 Db 680 CGGTGAGTCTGCGCAACCGCGCTGATGAGTCTGATCTGTGAGACGGTCTGG 739  
 Qy 233 GCTACTACCGGCGGGGGAACAC-----CTCCACCGCTGCGCGGAAGCG 277  
 Db 740 GCTACTACGCGCGGCGGCACTTACATAAAGCGCGGAGCAAGTCCCACTGCACG 799  
 Qy 278 TGAAGAGCTTCCCGAGCTTCCGAGCTTCCGAGGCTTCTGATCTCGGCTTTACA 337  
 Db 800 GCGGGGAATTCCTCCGAGCTTTGACGAAGTGGGCGCTACCGGCGTGGGGCTCAA 859  
 Qy 338 CCGGCGGCGGAGCTTCCATCGCTTCCGAGAGCGGAGTGGCGGCTGAGCGGAACG 397  
 Db 860 CGGCGGCGGAGTATTATCTCTTCCGCTGAGCAATATCCATTCTCGACGGAACG 919  
 Qy 398 TCCGGAAGGCTCTCCCGCTTCCGCGGGAAGCCCAAGAGAGAGCTTTTCG 457  
 Db 920 TGAAGCGGCTGCGCGCTGATGTCAGCGGCTGCGCGGGAAGAGAGTGG 979  
 Qy 458 CCTCGCCAGAGGCTCTCCCGAGGAGCTGACCCGAGGAGTGG-----A 505  
 Db 980 AAAAAGCGCTGTGGAGATCAGCAAGAGTCACTCCGCGAGAGGCGCTCAGAGCTTCA 1039  
 Qy 506 ACCAGGCTCTGAGAGCTGAGGAGCAAGGCTCTGCGGAACGCGCGCTTGGCGGG 565  
 Db 1040 ACCAGGAGATGATGATCTCGGGGCAATGATTGACCCCGCTCGAAGCGGAGTGGAGC 1099  
 Qy 566 CTTGCGCCCTAGAGGCTTCTGCGGAGGAGAGAGCGCC 604  
 Db 1100 TGTGTCCGTGAGCAAGCGCTGCTGCGCTTACGGAAC 1138

RESULT 5  
 US-09-252-991A-6783  
 ; Sequence 6783, Application US/09252991A

Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubinfeld et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196,136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 6783  
 ; LENGTH: 549  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-6783

Query Match 11.0%; Score 107.2; DB 4; Length 549;  
 Best Local Similarity 54.4%; Pred. No. 7.6e-13;  
 Matches 248; Conservative 0; Mismatches 193; Indels 15; Gaps 1;

Qy 55 CTCCCTGCGGGGAGAGAACCTTACCGGCTCTGCGAGGCTCTTCTGAG 114  
 Db 60 CTGCTCTGCGAGCAGGCGATACCCCTTACCGGCTGAGGCTCTGGAATCATGCTGAG 119  
 Qy 115 CAGACCCGGGTGAGAGAGCCCTCCCTATTACCGCGCTTCTGAGGCTTTCCACC 174  
 Db 120 CAGACCCAGGTACACACCGTCTCGTTACTTCAACGCTTCAATGAGCGCTTGGCCGAC 179  
 Qy 175 CTGAAGGCTCTGCGGCGCTTCTGAGAGGCTCTTGAAGGCTCTGAGGCTGAGGCGCG 234  
 Db 180 GTCAGGACACTGCGCGCGCGCGCGCGAGAGCAAGTCTGACCTGTGAGCGGCTGCG 239  
 Qy 235 TACTACCGGCGGAGAACACTTCCACCGCTGCGCGGAAGCT-----G 279  
 Db 240 TACTACAGCGTGGCGGCACTGACAAAGCCGCAAGTCTGTGAGCGGATGCG 299  
 Qy 280 GAGGAGCTTCCCGAGCTTCCGAGCTTCCGAGGCTTCTGAGCTTTACACC 339  
 Db 300 GGGAGTTCCTCCCGAGCTGAGCAACTGCGCGAGCTGCGCGATCGCGCTCCACC 359  
 Qy 340 GCGGCGGCTGCTCTCATGCTTGGGAGCGGAGTGGGCGGTGAGAGGAGCTC 399  
 Db 360 GCTGAGCACTGCGCACTGCTGATGAGGCTTGGCGAGCAGATCTGAGGAGCACTC 419  
 Qy 400 CGGAGGCTCTCCCGCTTCCGCGGGAAGCCCAAGAGAGAGCTTTCCGC 459  
 Db 420 AAGCGGCTGCGCGCTTACCTGCGGAGAGCGCTATCCCGGAGACGAGGTGCC 479  
 Qy 460 CTCGCCAGGCTCTTCCCGAGGAGGCGTGAACCG 495  
 Db 480 AGGCGCTGTGGAAGCGCGGAGGCTTACCCCG 515

RESULT 6  
 US-08-813-574-1  
 ; Sequence 1, Application US/08813574  
 ; Patent No. 6013473  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wei, Ying-Fei  
 ; TITLE OF INVENTION: Human Mucy  
 ; NUMBER OF SEQUENCES: 18  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SmithKline Beecham Corporation  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19406  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette

```
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/813,574
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/013,132
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1858 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-813-574-1
```

```
Query Match      10.7%; Score 104; DB 3; Length 1858;
Best Local Similarity 54.0%; Pred. No. 3,6e-12;
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;
```

```
QY 69 GGAAGAAGACCTTACCGGCTCTGCTCCGAGGCTCTTTCACAGAACCCGGGTGA 128
DB 498 GGAAGAGCGGGCTATGCTGTGTGCTCAGAGTCTGACAGACCCAGGTTGC 557
QY 129 GCAGGCGCTCCCTATTACCGCGCTTTTGAGCGCTTCCACCGTGAAGCGCTGAG 188
DB 558 CACTGTGATCACTACTATACCGGATGATGAGAAAGTGGCTCACTGAGAGACTGGC 617
QY 189 CCGGCTTCTCTGGAAGAGTCTTAGGGTCTGCAAGGGGCGGCGTACTACCGCGGAG 248
DB 618 CAGTGTCTCTCGAGAGAGTGAATCACTGTGGCTGGCTGCTACTATTCTGTGG 677
QY 249 GGAACCTTCCACCGGCTGGCGGAGCGTGAAGAGCTTCCCGAGCTT----- 300
DB 678 CCGGCGGCTGCAAGAGGAGCTCGAAGGTGTAGAGAGCTAGGGGCGCATGCGCAG 737
QY 301 -----GCCGAGCTTCGAGGAGCTTCTGCTCGGAGCTTACACCGCGCGCGGT 350
DB 738 TACAGAGAGACCTCTGACAGAGTCTGCTGCGCGTGGGCGCTACACAGCTGGGCTAT 797
QY 351 GGCCTTCATCGCTTCTGAGAGCGGAGCGGCGGTGAACGGGAACGTCCGAGGGTCTT 410
DB 798 TGCCTCTATCGCTTGTGGCCAGGCAACCGGTGTGTGATGCAAGTACAGCGGTGCT 857
QY 411 CTCGCGCTCTTTCGCGCC--GGGAAGCCCAAGAGAAAGAGTCTTCCGCTCGCCGAG 468
DB 858 GTGCCGTGTCCAGCATTTGTGTGCTATCCACAGACACCTTGTTCACAGAGCTGTG 917
QY 469 GGCCTCTCTCCGAGGCGGTGACCGG-----GGGAGTGTGAACACGAGCCCTCAT 518
DB 918 GGGGTAGCCACAGAGCTGTGTGAGACCCAGCCGCGCAGAGATTCAACAGAGAGCCAT 977
QY 519 GGAAGCTCGGGGCGACGCTGTGCTGCGGAACGAGCGCCCTTGGGGGCGCTGCCCCCTAG 578
DB 978 GGAGCTAGGGGCGACAGTGTGTATCCCAAGAGGCCCACTGTGACAGCACTGTGTGTA 1037
QY 579 GGCCTTCTGCGGG 592
DB 1038 GAGCTGTGCGGG 1051
```

RESULT 7

```
US-09-651-656-38
Sequence 38, Application US/09651656
Patent No. 6340566
GENERAL INFORMATION:
APPLICANT: MCCUTHEN-MALONEY, SANDRA
TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,
FILE REFERENCE: IL-10689
CURRENT APPLICATION NUMBER: US/09/651,656
PRIOR FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 60/192,764
NUMBER OF SEQ ID NOS: 106
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 2293
TYPE: DNA
ORGANISM: Escherichia coli
US-09-651-656-38
```

```
Query Match      8.9%; Score 87.2; DB 4; Length 2293;
Best Local Similarity 54.1%; Pred. No. 7.2e-09;
Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;
```

```
QY 53 CCTTCCTTGGCGGGGAGAGACCTTACCGGCTCTGCTCGAGGCTCTTCTGC 112
DB 1161 CTCTCCCTCGCAATTGACAAAGCCCTCAAAAGTATGCTCAGAAATGATGTTGC 1220
QY 113 AGCAGACCCGGGTGAGAGCGCCCTCCCTATTACCGCGCTTTCGAGGCGTTTCCCA 172
DB 1221 AACCAACTCAGGTGCGACCTGTATCCCTTATTGAACGCTTATGCGCGCTTCCGA 1280
QY 173 CCTTGAAGCCCTGCGCGGCTTCCCTGGAAGAGTCTTAGGCTGCGAGGCGCGG 232
DB 1281 CCGTACCGGATCTCGCCAAATCGCGCTCGACGAAGTCTCACTGTGTGACACCGGCTTG 1340
QY 233 GCTACTACCGGCGGCGGAACCTTCACCGCCTGGCGCGGAGCGTGAAG----- 282
DB 1341 GCTATTACCGCGCGCGCAATCTGCAATAAGCGGCAACAAGTGGCGACCTTACAG 1400
QY 283 -----GAGCTTCCCGGAGCTTTCGCGAGCTTTCGGGGGCTTCTGCTCGGGCTTACA 337
DB 1401 GCGGTAATTCGGAACCTTTGAGAAATTGACACTCGCGGCTCGGGGCTTCCA 1460
QY 338 CCGCGCGCGGAGTGTCTCATCGCTTTCGGGAGCGGAGTGGCGGCGGTGAACGGGAACG 397
DB 1461 CCGCAGGCGCATTTCTCGCTTCTGTGGTTAAGCACTTCCGATTCTGACCGGTAAAG 1520
QY 398 TCCGAGGGTCTCTCCGCTCTTTCG 425
DB 1521 TCAACGCGGTGCGCGCTGCTATGC 1548
```

```
RESULT 8
US-09-650-855-38
Sequence 38, Application US/09650855
Patent No. 6365355
GENERAL INFORMATION:
APPLICANT: MCCUTHEN-MALONEY, SANDRA
TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA
MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA
FILE REFERENCE: IL-10284
CURRENT APPLICATION NUMBER: US/09/650,855
PRIOR FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 60/192,764
NUMBER OF SEQ ID NOS: 106
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
```

LENGTH: 2293  
 TYPE: DNA  
 ORGANISM: Escherichia coli  
 US-09-650-855-38

Query Match  
 Best Local Similarity 54.1%; Pred. No. 7.2e-09;  
 Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;

53 CCTCTCCCTGGCGGGGAGAGACCTTACCGCGTCTGCTCCGAGGCTCTTCTGC 112  
 1161 CTCTGCGCTGGCAATTTGACAAAGCGCCCTACAAAGTATGCTCTCAGAAATGATGTTGC 1220  
 113 AGCAGACCCGGGTGGAGAGCCCTCTCTTATACCGCGCTTTCTGAGAGCTTTCCCA 172  
 1221 AACAACTCAGGTGGCGACCGTTATCCCTTATTTGAAGCTTTATAGCGCGCTTCCGA 1280  
 173 CCTGAAGAGCCCTGGCGCGGCTTCCCTGGAAGAGTCTTAAAGGCTGGCAGGGGCGG 232  
 1281 CGGTGACCGATCTGCGCAATGCGCGCTCGACGAAGTTCTCACTTGTGACCGGCTTG 1340  
 233 GCTACTACCGCGCGGCGGACACCTTCAACCGCTGGCGCGAAGCGTGGAG----- 282  
 1341 GCTATTAAGCGCGCGCGCAATCTGATTAAGCGGCAACAAGTGGCGACTTACAG 1400  
 283 -----GAGCTTCCCCAGCTTCCCGGAGCTTGGGGGCTTCTGCTTGGGCTTACA 337  
 1401 GCGGTAAATTCCTCGAACAATTGAGAGATTGACAGCTCCGCGCGCTCGGGGCTTCCA 1460  
 338 CCGCGCGCGCGGCGGCTTCCATCGCTTGGGAGAGCGGCGGCTGGAGCGGGAAG 397  
 1461 CCGGAGCGCGCATTTCTCTGCTTTCTGCGTAAACACTTCCGATTTCCAGCGTACG 1520  
 398 TCCGAGGCTCTCTCCGCGCTTTCG 425  
 1521 TCAAACGGGTCTGGCGCGCTGCTATGC 1548

RESULT 9  
 US-09-103-840A-2  
 ; Sequence 2, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; FILE REFERENCE: 24366-20007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103.840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 4403765  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; FEATURE:  
 ; OTHER INFORMATION: CDC 1551  
 ; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
 ; OTHER INFORMATION: represent a, t, c or g  
 ; US-09-103-840A-2

Query Match  
 Best Local Similarity 8.9%; Score 86.6; DB 3; Length 4403765;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;  
 22 CTCTGCGCTGTACCGGAAAGCGCGCCCTCTCCCTGGCGGGG--GAGAGAGAC 78  
 4022631 CTCTGCGCTGTATCAGCGATTCGCGGATCTGCGCGGAGAGCGCGGCTGACG 4022690  
 79 CTTTACCGCGTCTGTCTCGAGGTCTTCTGACAGACCGCGGTGAGACAGGCCCTC 138

Db 4022691 CCGTGGCAGATCTGTCTAGGAGATTCACTGACAGAGAGCGCGCGCGCGGTGCTG 4022750  
 139 CCTATTATACCGCGCTTTCTGAGCGCTTTCCACACCTGAAGGCCCTGCGCGGCTTCC 198  
 4022751 GCGATCTGGCCGAGCTGGGTGCGCGCGCGCCACGCGGTGCGGACCGCGCCAGC 4022810  
 199 CTGAAAGGTCCTTAGGTGTGCAAGGGCGGGCTACTACCGCGCGGAGAACCTC 258  
 4022811 ACCCGCATGTGTAACGGCTGGGAGCAAGCTGGGCTATCCAGCGAGCCAGTCTTA 4022870  
 259 CACCGCTGGCGCGCAAGCGT-----GAGAGCTTCCCCGAGCTTCC 303  
 4022871 CACGATGCGCCACGTCATGATCGCCCGGACCAACATACGATGTGCGCGAGATATGAG 4022930  
 304 GAGCTTGGGGGCTTCTGTGCTGTGGCTTTACACCGCGCGGCGGTGCTTCAATGCC 363  
 4022931 ATCTGTGTCACCTTGGCGGGGCTGGGAGCTACACCGCGCGCGGTGGCTGTTCGCT 4022990  
 364 TTGGGAGAGCGGTGGCGGCGGTGAGAGGAACTCCGAGAGGTCTCTCCGCTCTT- 422  
 4022991 TACCGCAGCGGCGGTGGGTGAGACCAATGTGCGCGCGGTGGTGGCGCGCTT 4023050  
 423 -----CGCGCGGAAAGCCCAAGAGAGAGCTTTTGGCCCTCGCCAGGAGC 471  
 4023051 CACGCGCGCGCGGACCGCGGTGCGCATGCTGCGCGCGGACACAGCGCATCTTGGCG 4023110  
 472 CTCTCTCCGAGGCGGTGAGACCGGAGGTGTGAACAAGGCTTATGAGCTTGGGCGC 531  
 4023111 CTGTGCGCGACCGCGAGACCGGCTGTAATTTTCGTCGCGCTGATGAGTGGGTGCG 4023170  
 532 ACGGTCTGCTGCGGAAAGCGCGCGGTGGGCGCTGCTTGAAGGCGCTTTCGCGG 591  
 4023171 ACGGTGTGACACCGCGCGGACCGCGGTGCGGTATGCTGCGTGAAGTGTGCGATGG 4023230  
 592 GGAAGAGAGCGCGCGGCGCTTACCGCGCGCGGAGAGCGCGGCGGAGAGAGCGC 651  
 4023231 CGCATGCGGTTATCCGCGCTCGGAGCGTCC---GCCGCGCGGGGCAAGCCTACACC 4023287  
 652 CTCTGCGCGCTGCTGCTCTCTGCGGCGG 678  
 4023288 GGAACCGACCGCGCAATCTCGCGGACGG 4023314

RESULT 10  
 US-09-103-840A-1  
 ; Sequence 1, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; FILE REFERENCE: 24366-20007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103.840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 4411529  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; OTHER INFORMATION: H37Rv  
 ; US-09-103-840A-1

Query Match  
 Best Local Similarity 8.9%; Score 86.6; DB 3; Length 4411529;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;  
 22 CTCTGCGCTGTACCGGAAAGCGCGCCCTCTCCCTGGCGGGG--GAGAGAGAC 78

Db 4030550 CTTCTGCTTGATCAGGATCCGACCGGGATCTGCCCTGGCCAGAGCCCGGTCTCAGC 4030609  
Qy 79 CCTTACCGGCTCTGCTCTCCGAGGTCTTCTGCAGACACCCGGGTGGAGCAGCCCTC 138  
Db 4030610 CCGTGGCAGATCTCGTCAAGGATTCATGCTGCAGACAGCCGGCCGGGTCTG 4030669  
Qy 139 CCTATTAACCGCGCTTCTGAGAGGCTTTCCACCCCTGAAGGGCCCTGGCCGGCTTC 198  
Db 4030670 GCGATCTGGCCGACTGGGTGCGCGGTGGCCACCGCTGGCCACCGCCAGCAG 4030729  
Qy 199 CTGGAAGAGGTCTTAAAGGTCTGCGAGGGGCGGCTTACTACCGCGGGCGGAACACTC 258  
Db 4030730 ACCGCCAGATGTATTAACGCGCTGGGGCAAGCTGGGCTATCCAGGCGCAACGGCTTA 4030789  
Qy 259 CACCGCCCTGGCCCAAGGCTG-----GAGAGCTTCCCGCAGCTTGGCC 303  
Db 4030790 CACGAGTCCGACCCGTATGCGCCGACCAACATACGCTGGCCGACGATATCAG 4030849  
Qy 304 GAGCTTCGGGGCTTCTGCTGCTGCGGCTTACACCGCGCGCGGTGGCTCCATCGCC 363  
Db 4030850 ATCTGTGATCACTTGGCGGGGCTGGGAGCTACACCGCGCGGTGGCTTTCCT 4030909  
Qy 364 TTCCGGAGAGCGGTGGCGCGGTGAGCGGAACTCCGAGAGGTCTCTCCGCTCTT 422  
Db 4030910 TACCGCAGCGGGTGGCGGTGTGACCAATGTGCGCGGTGGCTCCGCGCTT 4030969  
Qy 423 -----CGCCGGGAAAGCCCAAGAGAGAGCTTTTCCGCTCCGCGCGAGGC 471  
Db 4030970 CACGGCGCGCGGACCGCGGTGCGCTCATCGTCCGCGCAACAGCTGTTGGCG 4031029  
Qy 472 CTCTCTCCCGAGGCGGTGACCGCGGGGTGTGAACCAAGGCTCATGAGCTCGGGCC 531  
Db 4031030 CTGTTGCGGACCGCGAGAGCGGCTGAATTTTCGCTCGCTGATGAGTTGGGTGG 4031089  
Qy 532 ACCGCTGCTGCGGAAAGCGCCCGTGGGGGCGCTGCGCCCTTGAAGGCTTCTCGCG 591  
Db 4031090 ACCGTGTGACCGCGCCCAACCGCGGTGCGGTATGCGCTGAGCTGGTCCGATG 4031149  
Qy 592 GGGAGAGAGGCGCCCGGCGCTACCCCGCGCCAGAGAGCGCGGCGGAAGAGAGCGG 651  
Db 4031150 CGGCATCGCGGTATTCGCGCGTGGAGCGTCTC---GCCGCGCGGGGAGGCTTACAC 4031206  
Qy 652 CTGCTGCGCTCTGCTCTCTCTCGGCGG 678  
Db 4031207 GGAACGACCGCCCAAGTCCGCGAGCG 4031233

RESULT 11  
US-08-961-527-12/C  
Sequence 12, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunesh  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9909 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-12

Query Match 8.4%; Score 82; DB 4; Length 9909;  
Best Local Similarity 52.2%; Pred. No. 8.6e-08;  
Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

Qy 22 CTCTCGCTGTACCGGAAACCGCCCGCTCCCTGGCGGGGAGAGAGACCT 81  
Db 5808 CTTCTGCTGGTATGATGAAACAAAGAGATTTCCTTGAGAGAGAAATCT 5749  
Qy 82 TACCGGCTCTGTCTCCGAGGTCTTTCAGACAGACCGGGTGGAGAGCCCTCCC 141  
Db 5748 TATCATCTGGATATCTGAATATGCTTACAGACACAGGGTGTATACATATCT 5689  
Qy 142 TATTACCGCGCTTCTGAGAGCGCTTCCACCTGAAAGCGCTGCGCGGCTTCCCTG 201  
Db 5688 TACTAGAAAGATTTTGTGACTGTTTCCAACTGTGCAAACTGCGCACTGCGCTGAG 5629  
Qy 202 GAAGAGTCTTGAAGGTGTGCGAGGGGCGGCTA-----CTACCGCGG 246  
Db 5628 GAGAGTACTGAAGAGCTTGGAGGGGCTTGGCTATATTTCTGAGTTCCCAATATCAG 5569  
Qy 247 GCGGAACACTTCCACCGCGCTGCGGAGAGGTGAGAGAGCTTCCCGAGCTTCCGAG 306  
Db 5568 GCTGACGCCGACGATTTATGACTGACTTGTGGCCAAATTTCCAAATACCTTGAAGA 5509  
Qy 307 CTGGGGGCTTCTGTGCTTGGGCTTACACCGCGCGCGGTGCGCTCATCGCCTTC 366  
Db 5508 ATTTCCAGCTTGAAGAGGATTTGACCTTACACAGCAGACCAATTCAGATTTGCTTT 5449  
Qy 367 GGGAGCGGTGGGGGGGTGAGACGAGAGGTCTCTCTCCGCTTTG 424  
Db 5448 AACTTGCTGAGCAGCTGTATGATGTATGTATGTATGCGGCTTGGCGCTGTTTG 5391

RESULT 12  
US-09-107-532A-3654  
Sequence 3654, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998



```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arianello, Pamela Deneke
REGISTRATION NUMBER: 40,489
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3654:
SEQUENCE CHARACTERISTICS:
LENGTH: 1212 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (8) LOCATION 1...1212
SEQUENCE DESCRIPTION: SEQ ID NO: 3654:
US-09-107-532A-3654

Query Match      6.5%; Score 63.6; DB 4; Length 1212;
Best Local Similarity 47.6%; Pred. No. 0.00029;
Matches 278; Conservative 0; Mismatches 279; Indels 27; Gaps 2;

Qy 31 TGGTACCGGGAAGCGCCGCCCTCCCTGCGGGGGGAGAGACCTTACCGCGCTC 90
Db 100 TGGTATGACAGAGAAAGAAACCTGCTGGGCGCTATATCGTGAATCTTATCGTATC 159
Qy 91 CTGCTCTCCGAGGCTCTTTCAGAGAGACCGGGTGAAGAGCCCTCCCTATTAACGC 150
Db 160 TGGATCTCTGAGATCATGCTTCAGCAAAACAGAGTAGACACAGTGAATGATTTAT 219
Qy 151 GCGTTCTGAGGCGCTTCCACCGCTGAAGCCCTGCGCGGCGCTCCCTGGAAGAGTGC 210
Db 220 CGTTTCATGGAATGGTTTCTCAATGGAAGAGCTAGCAAAATGCTCCAGAAAGAGCTT 279
Qy 211 CTTAGGCTCTGCGAGCGGCGGCTACTACCGCGCGGCGAACAACCTC----- 258
Db 280 TTAAAGATGAGGAGGCGCTTGTATTACTCAAGAGCCGGAACATCAAGCTGTGCT 339
Qy 259 ---CACGGCTGGCCCGAAGCGTGAAGAGCTTCCCGAGCTTCGCGAGCTTCGGGGG 315
Db 340 AAACAGATCATGTGAGAAATTGATGGGGAATGCGCGAAACGCTGAAGAAATTAGTCA 399
Qy 316 CTTCCTGTCTCGGCGCTTACACCGCGCGCGGTGCTCATGCTTCCTTGGGGAGCGG 375
Db 400 TTGAAGGAATAGAGCAATATACAGACAGCGCGATGCAATGCAATTCGATTGCACTTCA 459
Qy 376 GTGCGCGCGTGAAGCGGAAGCGTCCGAGGGTCTCTCCGCTCTTCCGCGGGAAGC 435
Db 460 GAACCTGCAAGTTGACGCAAGTGAATGAGTGAATGTTTGGATTTGATTTGCAAGCA 519
Qy 436 CCCAAGAGAA-----GAGCTTTTGCCTTCGCGCAAGGCGCTCTCCCGAG 483
Db 520 GACATCCCAAAAGCTTCTAGCAGAGAAATCTTTGATGAAAGGATGCGGAATATCATGAT 579
Qy 484 GGCCTGACCCCGGGGTGTGAACCAAGCCCTCATGAGCTTCGGGGCAAGGTCTGCTG 543
Db 580 GAAAGACATCCAGGTGAATTCACCAACCAATGATGATCTAGGTTCACTATCTGTACA 639
Qy 544 CCGAAGCGCCCGCTTGGCGGGCGCTCCCTAGGGGCGCTCTG 587
Db 640 CCAACTTCTCCAAAGTGAAGCTTGTCCGATCCAGGCTTTTG 683
```

```

RESULT 13
US-09-134-000C-1710
Sequence 1710, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1710
LENGTH: 1188
TYPE: DNA
ORGANISM: Enterococcus faecalis
US-09-134-000C-1710
```

```

Query Match      6.4%; Score 62.4; DB 4; Length 1188;
Best Local Similarity 48.8%; Pred. No. 0.0005;
Matches 211; Conservative 0; Mismatches 206; Indels 15; Gaps 1;

Qy 8 CTTGGCGGAAGCCCTCTGCTGTACCGGGAAGCGCCGCCCTCCCTGGCGGG 67
Db 53 CATTCAGAAAGATTTTATGCTGTATGACGAGAAACGCAATTTACCTTGGCGAG 112
Qy 68 GGGAGAGAGACCTTACCGGCTCTGCTCCGAGGTCCTTCTCAGAGACCGGGTGG 127
Db 113 GGAATACAGATGATATGATTTGATTTCTGAATTAATGCTTCAACAAATCTCGTAG 172
Qy 128 AGCAGCCCTCCCTATTTACCGCGCTTCTGAGACGCTTCCACCTGAAGGCGCTGG 187
Db 173 ATACGATCATGATTTATTTATGATTTAATGAAATGATTTCCAGCATTTCAAGATTAG 232
Qy 188 CCGGCGCTTCCCTGGAAGAGTCTTAAGGTCTGCGAGGGGCGGCTACTACCGGCGGG 247
Db 233 CGGAAGCGCGAGATGAATGATTTGTAAGCTTGGGAAGGCTTAGTTACTATTCACGAG 292
Qy 248 C-----GGAACACTCCACCGCGCTGCGCGCAAGGTGAAGAGCTTCCC 292
Db 293 GCGTAATTAAAGTGGCAGCGCAACGATTTTCAAGATTTGTTGGGAAATGCTTG 352
Qy 293 CGAGCTTCGCGAGCTTCGGGGGCTTCTGCTCTGCGGCTTACACCGCGCGCGGTGG 352
Db 353 ACACATCGAAGATATTTCGAGTTTAAAGAAATCGGCCCTTATACGCTGTGCGATTG 412
Qy 353 CTTCCATCGCTTGGGGAAGCGGTGCGCGGCTGACGGAACGTCCGAGGGTCTCT 412
Db 413 GCAATATTGCTTTAATCTCCGCAACAGCAATTTGACGATGATGCGGTGATGTA 472
Qy 413 CCGGCTCTTGG 424
Db 473 GCCGTTATTGG 484

RESULT 14
US-09-385-028-23
Sequence 23, Application US/09385028
Patent No. 6232106
GENERAL INFORMATION:
APPLICANT: Susan E. Jensen
APPLICANT: Kwamena A. Aidoo
APPLICANT: Ashish S. Paraskar
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
Patent No. 6232106
TITLE OF INVENTION: Acid Biosynthesis
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
```

STREET: The Jenifer Building, 400 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/385,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/790,462  
FILING DATE: 29-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: D. Douglas Price  
REGISTRATION NUMBER: 24,514  
REFERENCE/DOCKET NUMBER: 1418/P57452US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-6666  
TELEFAX: (202) 39305350  
TELEX: RCA 248593 IDEA UR  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1227 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-385-028-23

Query Match 6.4%; Score 62.2; DB 3; Length 1227;  
Best Local Similarity 44.4%; Pred. No. 0.00055;  
Matches 345; Conservative 0; Mismatches 423; Indels 9; Gaps 2;

157 CTGGAAGCGCTTCCACCTGAAAGCCCTGGCCGCGCTTCCCTGGAAGAGTCTTAGG 216  
88 CTGGCCGGGCTGGGTCCAGAAAGCCGAGCCGGGTGACGCTGTGGACAGCCAG 147  
217 GTCTGGCAAGGGGGGGGCTACTACCGCGGGGGGGAACACTCCACCGCTGGCCGGAAG 276  
148 GTGTGGCTGTGACCTTGACCGCGGGGGGGCGGCGCTCTGGGCGAACCGCGCTTACCC 207  
277 GTGGAAGAGCTTCCCGGAGCTTCGCGAGCTTCGCGGGCTTCTGTGCTCGGGCCCTTAC 336  
208 GCGGTACAGAGCGCGCGCGCTTCCGATGTGACCCGCACTCCGACCTGGTGCGCGCC 267  
337 ACCGCGCGCGGGGTGCTTCATCGCTTCGGGGAGCGGGGTGGCGCGGTGACGG---- 392  
268 AACCCGGAATCGGCGTGTTCATCCGATGAGACCCGAGCACTCCGCGCTGCGCTCG 327  
393 --GAAGCTCGGAGGGTCTCTCCCGCGCTTTCGCGCGGGAAGCCCGAAGAGAGAG 450  
328 ATGCTACCCGGAGCTTCTCTGCGCGCGCGCGAGCGCTGCGCGCGCGGTGCGGAG 387  
451 CTTTTCGCTCCCGAGGCTCTCTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 510  
388 CTGCTGAGAGATCTCTGGCGGGGCTGTGAAGGGGAGGCGGCTGCACTGTGCGCC 447  
511 GCCCTCATGAGTCTGGGGGCAAGCTGTGCTGCGCAAAAGCGCGCTTGGCGGGCTTGC 570  
448 GGACTACGATCCCGGTGCTCTGCGGGGTATCACTCTTTCGCGCGCGGTGACGAG 507  
571 CCCCTAGGGGCTTCTGCGCGGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 630  
508 CGCGGGAATTCATGAGAGAGCGCAGCGGCTCTTATGACCGCGGCTTACACCCGAG 567  
631 CGCGGAGAGAGAGAGCGCTGTGCGCTGCTCTCTCTCGGCGGAGAGGGGAGTGCAC 690  
568 CAGGTGCGCAAGGCGCGGAGCAACTGACGCTATCTGCGGAGGCTGTGTGAGGAGCGG 627

691 CTGGAAGCTTGAAGGGCGCTTCCAGGCGCTCTAAGCGGCGTCCCGCTTTCCTCCCTGAG 750  
628 ATCGAAGAACCCGGGACCGACCTGATACGCGGCTCTCATACACAGGTGCGGCGGG 687  
751 GAGCTCCCGGGCGGAGGCGGCTTTCGAGGTGAGTCTAAGCCCTTGAAGGAGTGGCC 810  
688 CATCTGCGGCTCAGAGAGATGTCCGATGTCCGGCTGCTGTGTGCGCGGTACCGGC 747  
811 CAGCCCTCACCACCGGAGCTTTCGCTGAGGTGCGGGGAGCCCTTTGGGAAGGGAG 870  
748 ACCACACACAGCCAGCGAGCTGAGCTGCTCAGCTCTCAGCCCTGACCGAGCCGAGCTG 807  
871 GGGAGAGACCCCTTGAAGAGGCC--CTACCCAGCTCATGAGAGGTGCTCCGC 924  
808 GGGGCTCACCAGAGACCGGCGCTGCTGCGCAAGCGGTGAGAGAGCTGCTCGC 864

RESULT 15  
US-09-726-614-23  
Sequence 23, Application US/09726614  
Patent No. 6514735

GENERAL INFORMATION:  
APPLICANT: Susan E. Jensen  
APPLICANT: Kwamena A Aidoo  
APPLICANT: Aehish S. Parthkar  
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
Patent No. 6514735  
TITLE OF INVENTION: Acid Biosynthesis  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESS: JACOBSON, PRICE, HOLMAN & STERN, PLLC  
STREET: The Jenifer Building, 400 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/726,614  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/790,462  
FILING DATE: 29-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: D. Douglas Price  
REGISTRATION NUMBER: 24,514  
REFERENCE/DOCKET NUMBER: 1418/P57452US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-6666  
TELEFAX: (202) 39305350  
TELEX: RCA 248593 IDEA UR  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1227 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-726-614-23

Query Match 6.4%; Score 62.2; DB 4; Length 1227;  
Best Local Similarity 44.4%; Pred. No. 0.00055;  
Matches 345; Conservative 0; Mismatches 423; Indels 9; Gaps 2;

157 CTGGAAGCGCTTCCACCTGAAAGCCCTGGCCGCGCTTCCCTGGAAGAGTCTTAGG 216  
88 CTGGCCGGGCTGGGTCCAGAAAGCCGAGCCGGGTGACGCTGTGGAGACGCGAGCCAG 147

```

QY 217 GTCTGCGAGGGGGGGGGCTTACCGGCGGGCGGAACACCTCAACCGCTTGGCCGAAGC 276
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148 GTGTGGCTGTGTGACTCTGCACGCGGGGCCCGGGCCCTCTGGGCGACCGCCCTTCAAC 207
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 277 GTGAGAGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCC 336
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 208 GCGGTGACGAGCGCGCCCGGCTTCCGATGTGACCGCGACCTCCCAACTGTGTGCGGCC 267
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 337 ACCGCGGCGGGGTGGCTTCCATGCTTCCGGGAGCGGGGTGGCGGGGTGACG---- 392
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 268 AACCGGAGTGGGCTTCCATCGCATGACGACCGCAGCCTCCCGCTGCGCTCG 327
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 393 --GAACGTCGAGAGGCTTCTCCGCTCTTGGCCCGGGAAGCCCAAGAGAGAG 450
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 328 ATGCTCACCCGGACTTCTGTGGCCCGCGCGCGAGGCGCTGCGCCCGCGGTGCGGAG 387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 451 CTTTTCGCTCGCCCGAGGCTCTCCCGAGGCGGTGACCCCGGGGGTGTGAAACAG 510
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 CTGTGTGACGAGATCTGTGGCGGGCTGTGTAAGGGGAGCGGCGGCTGCACTGTGCGC 447
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 511 GCCCTCATGAGCTCGGGGCGACGGTCTGCTGCGAAACGCGCCCGTTGCGGGGCTGC 570
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 448 GGACTGACGATCCCGGTGCCCTCGCGGGTATCACCTGTCTTCCGCGCGGTGACGAC 507
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 571 CCCCTAAGGGGCTTCTGCGGGGGAAGAGCCCGCGGCTAACCCCGGCTTCCGAGAG 630
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 508 CGCGGGAGTTTCATCGAGGACCGGAGCGGGTCTCATCGACCGGGCTACACCCGAG 567
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 631 CGCGGGCGAAGAGAGAGCGGCTCTGCGCCCTCTCTCTCGGGCGGAAAGGGGTGCAC 690
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 568 CAGGTGCGCAAGGCCCGGAGAGAACTGACGGCTATCTGCGGAGCTGTGTGAGAGCGG 627
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 691 CTGGAAGGCTTGAAGGGGCGCTTCCAGGGCTTACGAGCGTCCCTCTTCCCTTGA 750
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 628 ATCGAAGAACCGGGCACCGACTGATCAGCCGCTCGTCAATGACCAAGTGCGGCGGG 687
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 751 GAGCTTCCCGGGCGGGAGCGGCTTGGGGTGAAGTCTAGGCCCCCTAGGAGAGTGC 810
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 688 CATCTGCGGGTCAAGAGATGTGCCGATGTGCCGCTGTGTGTGCGGCTGACGCGC 747
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 811 CAGGCTCTCAACCGACCGAGGCTTCCGCTGAGGTGCGGGGGGCTTGGGGAAGGAG 870
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 748 ACCACCAACGAGCGGAGGCTGAGCTGTCAAGCTGTCAACCGAGCGGAGCTGCGC 807
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 871 GGGGAGGAGCCCTGGAAGAGGCC--CTACCAAGCTCATGAGAAAGTGTCTCGC 924
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 808 GGGGCTCAACCGAGGAGCCGCGCTGTGCTGCCAAGGGCGGTGAGGAGCTGTGCGC 864
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: March 4, 2004, 05:48:50  
 Job time : 97 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2004, 03:03:25 ; Search time 358 Seconds  
(without alignments)  
11569.813 Million cell updates/sec

Title: US-09-938-901-1

Perfect score: 975  
Sequence: 1 gtggaagccgcgcgaagc.....tagtccccccgcgaacca 975

Scoring table: IDENTITY NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: geneseq19808:\*  
2: geneseq19908:\*  
3: geneseq20008:\*  
4: geneseq2001as:\*  
5: geneseq2001bs:\*  
6: geneseq2002as:\*  
7: geneseq2003as:\*  
8: geneseq2003bs:\*  
9: geneseq2003cs:\*  
10: geneseq2004as:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	975	100.0	975	7	AB222143
2	104	10.7	1608	9	AB222143 Thermus t
3	104	10.7	1608	9	AB222143 Thermus t
4	104	10.7	1608	9	AB222143 Thermus t
5	104	10.7	1854	7	AB222143 Thermus t
6	104	10.7	1858	2	AB222143 Thermus t
7	102.6	10.5	1856	2	AB222143 Thermus t
8	102	10.5	1856	2	AB222143 Thermus t
9	95.6	9.8	349980	3	AA221610
10	95.4	9.8	110000	3	AA221610
11	92.8	9.5	528	6	ABK74292
12	89.2	9.1	1047	7	AB240024
13	87.2	8.9	2293	4	AA221610
14	86.6	8.9	912	7	AB279956
15	86.6	8.9	1312	7	AB279956
16	86.6	8.9	110000	4	AA199682_40
17	86.6	8.9	110000	4	AA199682_40
18	83.6	8.6	23532	3	AA221610
19	82	8.4	1173	7	ABX06945
20	82	8.4	9909	7	AAV52145
21	82	8.4	110000	7	AB222143
22	81.2	8.3	966	6	AA225855
23	81	8.3	960	2	AA225855

C	24	81	8.3	2077	2	AA296352
C	25	81	8.3	2077	2	AAV37388
C	26	80.4	8.2	1176	2	AA222143
C	27	80.2	8.2	1038	7	ACF69235
C	28	80.2	8.2	110000	7	ACF69235
C	29	80.2	8.2	243072	7	ACF69235
C	30	78.6	8.1	1206	8	ADBI0739
C	31	78.6	8.1	1206	8	ADBI0741
C	32	78.6	8.1	1206	8	ADBI0743
C	33	78.6	8.1	110000	8	ADBI0744
C	34	75	7.7	114955	2	AA222143
C	35	73.4	7.5	66788	4	AA222143
C	36	73.4	7.5	66788	7	ACF64444
C	37	72.6	7.4	349980	6	ABQ81847
C	38	71.8	7.4	1412	3	AA222143
C	39	71.8	7.4	1412	6	ABN22977
C	40	70.6	7.2	114955	2	AA222143
C	41	68.2	7.0	991	6	ABQ68775
C	42	68.2	7.0	3155	6	ABQ70783
C	43	66	6.8	35133	4	AB150991
C	44	66	6.8	35133	4	AB150990
C	45	65.6	6.7	2000	7	ADA71938

## ALIGNMENTS

RESULT 1	AB222143	standard; DNA; 975 BP.
XX	AB222143;	
XX	12-MAR-2003 (first entry)	
XX	Thermus thermophilus DNA repair enzyme MutY encoding DNA SEQ ID NO:1.	
XX	Thermus thermophilus; DNA repair enzyme; MutY; RecJ; TRCF;	
XX	biochemistry; molecular biology; research; gene; ds.	
XX	Thermus thermophilus.	
XX	Key	Location/Qualifiers
FT	CDS	1..975
FT		/*tag= a
FT		/partial
FT		/product= "MutY"
FT		/note= "DNA repair enzyme; no start or stop codons given"
XX	JP2002247985-A.	
XX	03-SEP-2002.	
XX	23-FEB-2001; 2001JP-00047762.	
XX	23-FEB-2001; 2001JP-00047762.	
XX	(RIKA) RIKAGAKU KENKYUSHO.	
XX	WPI: 2003-078924/08.	
XX	P-PSDB; ABP56413.	
XX	A DNA repair enzyme gene, a protein, a recombinant vector, a	
XX	transformant, preparation of DNA repair enzyme, repairing the error	
XX	sequence of a DNA, and prevention of error synthesis of a DNA sequence.	
XX	Claim 3; Page 13-14; 41pp; Japanese.	
XX	AB222143 to AB222146 encode the Thermus thermophilus DNA repair enzymes	
XX	MutY, RecJ, and TRCF given in ABP56413 to ABP56416. The enzymes can	
XX	be used as research reagents for biochemistry and molecular biology	
XX	Sequence 975 BP; 126 A; 347 C; 354 G; 148 T; 0 U; 0 Other;	

Query Match 100.0%; Score 975; DB 7; Length 975;  
 Best Local Similarity 100.0%; Pred. No. 4,5e-160;  
 Matches 975; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTGAGAGGCTGAGCGAAGAGCCCTCTGCTGTATCCGGGAAAACCGCGCCCTCCCG 60
DB 1 GTGAGAGGCTGAGCGAAGAGCCCTCTGCTGTATCCGGGAAAACCGCGCCCTCCCG 60
QY 61 TGGCGGGGAGAGAGACCTTACCGCGTCTGTGCTTCCGAGGCTCTTTCGACAGACC 120
DB 61 TGGCGGGGAGAGAGACCTTACCGCGTCTGTGCTTCCGAGGCTCTTTCGACAGACC 120
QY 121 CGGGTGGACAGAGCCCTCCCTTATACCGCGGCTTTCTGAGAGGCTTTCCACCTGAAG 180
DB 121 CGGGTGGACAGAGCCCTCCCTTATACCGCGGCTTTCTGAGAGGCTTTCCACCTGAAG 180
QY 181 GCGCTGGCGGAGCTTCCCTGAGAGGTCCTTAGAGGCTGAGAGGGGCGGAGCTACTAC 240
DB 181 GCGCTGGCGGAGCTTCCCTGAGAGGTCCTTAGAGGCTGAGAGGGGCGGAGCTACTAC 240
QY 241 CGGCGGGCGAACAACCTTCCAGCGCTGGCCGGAAGCGTGAAGAGCTTCCCGAGCTTC 300
DB 241 CGGCGGGCGAACAACCTTCCAGCGCTGGCCGGAAGCGTGAAGAGCTTCCCGAGCTTC 300
QY 301 GCGGAGCTTCGGGGGCTTCTGTGCTCGGCTTACACCGCGGCGGCGGAGGCTGCTGATC 360
DB 301 GCGGAGCTTCGGGGGCTTCTGTGCTCGGCTTACACCGCGGCGGCGGAGGCTGCTGATC 360
QY 361 GCGTTGAGGAGAGCGGTGAGCGGCGGTGAGAGGAGAGGTCGAGAGGCTCTTCCGCTC 420
DB 361 GCGTTGAGGAGAGCGGTGAGCGGCGGTGAGAGGAGAGGTCGAGAGGCTCTTCCGCTC 420
QY 421 TTGCGCCGAGAAAGCCCAAGAGAGAGCTTTTGCCCTGCGCCAGGAGCTTCTCCCG 480
DB 421 TTGCGCCGAGAAAGCCCAAGAGAGAGAGCTTTTGCCCTGCGCCAGGAGCTTCTCCCG 480
QY 481 GAGGGGCTGAGACCCGGGGGTGAGAACAGGCGCTATAGAGTCGGGGCCACAGGCTAC 540
DB 481 GAGGGGCTGAGACCCGGGGGTGAGAACAGGCGCTATAGAGTCGGGGCCACAGGCTAC 540
QY 541 CTGCGGAAAGCGCCCTTGTGCGGGGCTGCGCCCTGAGGGAGCTTGTGCGGGAGAGAG 600
DB 541 CTGCGGAAAGCGCCCTTGTGCGGGGCTGCGCCCTGAGGGAGCTTGTGCGGGAGAGAG 600
QY 601 GCGCCCGGAGCGCTACCCCGCGCCAGAGAGCGCGGCGAAGAGAGAGCGCTCTGCGCC 660
DB 601 GCGCCCGGAGCGCTACCCCGCGCCAGAGAGCGCGGCGAAGAGAGAGCGCTCTGCGCC 660
QY 661 CTGCTCTCTGCGGGGAGAGGGGGGTGACCTGAGAAAGGCTTGAAGGGGGCTTCCAGGGC 720
DB 661 CTGCTCTCTGCGGGGAGAGGGGGGTGACCTGAGAAAGGCTTGAAGGGGGCTTCCAGGGC 720
QY 721 CTCTACGAGCGTCCCTCTTTTCCCTGAGAGGCTTCCCGGGCGGAGAGCGGCTTCCGG 780
DB 721 CTCTACGAGCGTCCCTCTTTTCCCTGAGAGGCTTCCCGGGCGGAGAGCGGCTTCCGG 780
QY 781 GTGAGGTCTAGAGCCCTTAGGCGAGGTGCGCACGCGCTTACCCACCGAGGCTTGGCGTG 840
DB 781 GTGAGGTCTAGAGCCCTTAGGCGAGGTGCGCACGCGCTTACCCACCGAGGCTTGGCGTG 840
QY 841 GAGGTGCGGGGGGCGCTTGGAGAGGGGAGAGAGCCCTGAGAAAGGCGCCCTACCC 900
DB 841 GAGGTGCGGGGGGCGCTTGGAGAGGGGAGAGAGCCCTGAGAAAGGCGCCCTACCC 900
QY 901 AAGCTCATGAGAAAGGTGCTCGCAAGCGCTTCCCTCTGCTCATGCGGAGCGTATGTC 960
DB 901 AAGCTCATGAGAAAGGTGCTCGCAAGCGCTTCCCTCTGCTCATGCGGAGCGTATGTC 960
QY 961 CCCCTCCGAGCGCA 975
DB 961 CCCCTCCGAGCGCA 975

```

RESULT 2  
 ADE36258  
 ID ADE36258 standard; cDNA; 1608 BP.  
 AC ADE36258;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human MYH (hMYH) G1396T mutant cDNA.  
 XX  
 KW gene; ss; mutant; human; screening method; hMYH; base excision repair;  
 KW BER; APC; familial adenomatous polyposis; FAP;  
 KW multiple colorectal adenoma; carcinoma; bowel cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1608  
 FT /tag= a  
 FT /product= "Human MYH mutant protein (E466X)"  
 FT allele /replace(1396..9)  
 FT /tag= b  
 FT /note= "Base T in mutant replaces wild type base G"  
 XX  
 PN MO2003014390-A2.  
 XX  
 PD 20-FEB-2003.  
 XX  
 PF 02-AUG-2002; 2002MO-GB003591.  
 XX  
 PR 03-AUG-2001; 2001GB-00018995.  
 XX  
 PA (UYWA-) UNIV WALES COLLEGE OF MEDICINE.  
 XX  
 PI Sampson JR, Cheadle JP;  
 XX  
 DR MPI: 2003-256601/25.  
 DR P-PSDB; ADE36261.  
 XX  
 PT Screening, diagnostic and therapeutic methods in individuals with  
 PT predisposition towards having a cancer, such as colon cancer, using base  
 PT excision repair pathway or hMYH genes.  
 XX  
 PS Claim 9; SEQ ID NO 3; 66bp; English.  
 XX  
 CC This invention relates to a novel screening method for identifying an  
 CC individual having a predisposition towards a cancer. Specifically, it  
 CC refers to obtaining a test sample, preferably comprising the hMYH gene  
 CC that occurs in the base excision repair (BER) pathway, and comparing this  
 CC nucleic acid molecule to the corresponding region of the wild type  
 CC sequence. This BER pathway gene, hMYH, acts to protect against G:C to T:A  
 CC transverse mutations in a cancer marker gene such as APC that is seen in  
 CC familial adenomatous polyposis (FAP). As such, mutations identified in  
 CC hMYH are associated with the onset multiple colorectal adenomas and  
 CC carcinoma. The present invention describes a screening method for  
 CC individuals that works to identify differences comprising any one of  
 CC G382D, Y165C, E466X or Y904 variations in hMYH, this signifies a cancer  
 CC predisposition, particularly for bowel cancer.  
 CC sequence is the mutant G1396T hMYH cDNA that encodes a E466X mutation in  
 CC the encoded polypeptide in the invention.  
 XX  
 SQ Sequence 1608 BP; 337 A; 478 C; 502 G; 291 T; 0 U; 0 Other;

Query Match 10.7%; Score 104; DB 9; Length 1608;  
 Best Local Similarity 54.0%; Pred. No. 2.6e-09;  
 Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

```

QY 69 GAGAGAGACCTTACCGCGCTGCTGCTCCGAGGCTCTTTCGACGACGAGCCGGGTGA 128
DB 327 GAGACGAGGCGCATATGCTGTGTGAGGTCTCAGAGTATGCTGACGACGACGAGTTGC 386
QY 129 GAGGCGCTCCCTATTAACCGCGCTTCTGAGAGGCTTCCACCTGAAGGCCCTGAC 188

```

Db 387 CACTGTATCACTACTATACCGGATGATGACAAAGTGGCTTACATGCGAGACCTTGGC 446  
Qy 189 CGCGGCTTCCCTGGAGAGGTCTTAAAGGTGTGCGAGGGCGGGCTTACTACCGCGGGC 248  
Db 447 CAGTGTCTCCCTGGAGAGGTGAATCACTGTGGGCTGGCTGGGCTACTATTCTCGTGG 506  
Qy 249 GGAACACTCTCAACCGCTGGCGCCGAGGCTGAGAGAGCTTCCCGAGCTTC----- 300  
Db 507 CCGCGGCTGTGAGAGGAGCTCGAAGGTGTGTAGAGAGCTTAAAGGGGCCATGCGACG 566  
Qy 301 -----GCCGAGCTTGGGGGCTTCTGTGTGCGGCTTAAACCGCGGGCGGCT 350  
Db 567 TACAGCAGAGACCTTGACAGACTCTGCTGGCTGGGCGCTTACACAGCTGGGGCCAT 626  
Qy 351 GGCCTTCATCGCTTTCGGGAGCGGGTGTGCGCGGATGACCGGAACGTCCGAGGATCT 410  
Db 627 TGCTCTATGCGCTTTTGGCCAGGCAACGGGTGTGGATGGCAACGTAGCAGCGGTGCT 686  
Qy 411 CTCCCGCTTTCGCCC--GGAAAAGCCCCAAGAGAGAGCTTTTGGCCCTTGGCCAG 468  
Db 687 GTGCTGTGTCCGAGCCATGTGTGTGTATCCAGCAGCAGCCCTTGTTCCTCAGACTCTG 746  
Qy 469 GGCCTCTCCCGAGGGGCTGAGCCG-----GGGGTGTGAACGAGGCTTCAT 518  
Db 747 GGGTCTAGCCAGAGCTGTGTGACCCAGCCCGCCAGAGATTTCAACAGAGACCAT 806  
Qy 519 GAGAGCTGGGGCCACGGTCTGCTGCGGAAACGGCCCGTTCGGGGGCTGCCCCTAGG 578  
Db 807 GAGAGTAAAGGGCCACAGTGTGTATCCCAAGCGCCCACTGTGACAGCCAGTCCCTGTGA 866  
Qy 579 GGCCTTCTGCCGGG 592  
Db 867 GAGCTGTGCGCGG 880

RESULT 3  
ADE36259  
ID ADE36259 standard; cDNA; 1608 BP.

AC ADE36259;

DT 29-JAN-2004 (first entry)

DE Human MYH (hMYH) C270A mutant cDNA.

XX gene; ss: mutant; human; screening method; hMYH; base excision repair;  
KW BBR; APC; familial adenomatous polyposis; FAP;  
KW multiple colorectal adenoma; carcinoma; bowel cancer.

XX Homo sapiens.

OS location/Qualifiers

FT Key 1..1608  
FT CDS /\*tag= a  
FT allele /product= "Mutant MYH protein (Y90X)"  
FT /\*tag= b  
FT /note= "Base A in the mutant replaces wild type base C"

XX W02003014390-A2.

XX 20-FEB-2003.

XX 02-AUG-2002; 2002WO-GB003591.

XX 03-AUG-2001; 2001GB-00018995.

XX (UYMA-) UNIV WALES COLLEGE OF MEDICINE.

XX Sampson JR, Cheadle JP;

XX WPI; 2003-256601/25.

DR P-PSDB; ADE36262.

XX Screening, diagnostic and therapeutic methods in individuals with  
PT predilection towards having a cancer, such as colon cancer, using base  
PT excision repair pathway or hMYH genes.

PS Claim 10; SEQ ID NO 4; 66pp; English.

XX This invention relates to a novel screening method for identifying an  
CC individual having a predilection towards a cancer. Specifically, it  
CC refers to obtaining a test sample, preferably comprising the hMYH gene  
CC that occurs in the base excision repair (BER) pathway, and comparing this  
CC nucleic acid molecule to the corresponding region of the wild type  
CC sequence. This BER pathway gene, hMYH, acts to protect against G:C to T:A  
CC transverse mutations in a cancer marker gene such as APC that is seen in  
CC familial adenomatous polyposis (FAP). As such, mutations identified in  
CC hMYH are associated with the onset multiple colorectal adenomas and  
CC carcinoma. The present invention describes a screening method for  
CC individuals that works to identify differences comprising any one of  
CC G382D, Y165C, E466X or Y90X variations in hMYH, this signifies a cancer  
CC predilection, particularly for bowel cancer. This polynucleotide  
CC sequence is the mutant C270A hMYH cDNA that encodes a Y90X mutation in  
CC the encoded polypeptide of the invention.

XX Sequence 1608 BP; 338 A; 477 C; 503 G; 290 T; 0 U; 0 Other;

Query Match 10.7%; Score 104; DB 9; Length 1608;

Best Local Similarity 54.0%; Pred. No. 2.6e-09;  
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

Qy 69 GGAAGAAGACCTTACCGCGCTCTGCTCCGAGGCTCTTGTGAGAGACCGGGTGA 128  
Db 327 GAGCAGGCGGGCATATGCTGTGTGTCAAGGTATGTGACAGACCAAGTTTC 386  
Qy 129 GCAGGCTCTCCCTATTACCGCGCTTCTGAGCGCTTCCACCTGAAGGCTTGGC 188  
Db 387 CACTGTATCACTACTATACCGGATGATGACAAAGTGGCTTACATGCGAGACCTTGGC 446  
Qy 189 CGCGGCTTCCCTGGAGAGGTCTTAAAGGTGTGCGAGGGCGGGCTTACTACCGCGGGC 248  
Db 447 CAGTGTCTCCCTGGAGAGGTGAATCACTGTGGCTGGCTGTGCTACTATTCTCGTGG 506  
Qy 249 GGAACACTCTCAACCGCTGGCGCCGAGGCTGAGAGAGCTTCCCGAGCTTC----- 300  
Db 507 CCGCGGCTGTGAGAGGAGCTCGAAGGTGTGTAGAGAGCTTAAAGGGGCCATGCCACG 566  
Qy 301 -----GCCGAGCTTGGGGGCTTCTGTGTGCGGCTTAAACCGCGGGCGGCT 350  
Db 567 TACAGCAGAGACCTTGACAGACTCTGCTGGCTGGGCGCTTACACAGCTGGGGCCAT 626  
Qy 351 GGCCTTCATCGCTTTCGGGAGCGGGTGTGCGCGGATGACCGGAACGTCCGAGGATCT 410  
Db 627 TGCTCTATGCGCTTTTGGCCAGGCAACGGGTGTGGATGGCAACGTAGCAGCGGTGCT 686  
Qy 411 CTCCCGCTTTCGCCC--GGAAAAGCCCCAAGAGAGAGCTTTTGGCCCTTGGCCAG 468  
Db 687 GTGCTGTGTCCGAGCCATGTGTGTGTATCCAGCAGCAGCCCTTGTTCCTCAGACTCTG 746  
Qy 469 GGCCTCTCCCGAGGGGCTGAGCCG-----GGGGTGTGAACGAGGCTTCAT 518  
Db 747 GGGTCTAGCCAGAGCTGTGTGACCCAGCCCGCCAGAGATTTCAACAGAGACCAT 806  
Qy 519 GAGAGCTGGGGCCACGGTCTGCTGCGGAAACGGCCCGTTCGGGGGCTGCCCCTAGG 578  
Db 807 GAGAGTAAAGGGCCACAGTGTGTATCCCAAGCGCCCACTGTGACAGCCAGTCCCTGTGA 866  
Qy 579 GGCCTTCTGCCGGG 592  
Db 867 GAGCTGTGCGCGG 880

RESULT 4  
ABV72438

ID ABV72438 standard; DNA; 1854 BP.  
 AC ABV72438;  
 XX  
 DT 29-JAN-2003 (first entry)  
 XX  
 DE Nucleotide sequence of adenine mismatch glycosylase.  
 XX  
 KM Human; cancer; DNA repair enzyme; DNA damage preventing enzyme;  
 XX adenine mismatch glycosylase; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200276280-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-IL000231.  
 XX  
 PR 23-MAR-2001; 2001US-00815015.  
 PR 09-JUL-2001; 2001US-030338P.  
 XX  
 PA (YEDA ) YEDA RES & DEV CO LTD.  
 XX  
 PI Livneh Z, Paz-Elizur T, Blumenstein S;  
 DR WPI; 2003-023894/02.  
 XX  
 PT Determining the risk to develop cancer for prevention, early detection  
 PT and prognosis, comprises determining a level of parameter indicative of a  
 PT level of activity of a DNA repair/damage preventing enzyme in a tissue of  
 PT the subject.  
 XX  
 PS Disclosure; Page 86-87; 91pp; English.  
 XX  
 CC The specification describes a method for determining the risk of a  
 CC subject to develop cancer. The method comprises determining a level of  
 CC parameter indicative of a level of activity of a DNA repair/damage  
 CC preventing enzyme in a tissue of the subject, and determining the risk of  
 CC the subject to develop the cancer according to the level. The method is  
 CC useful for determining risk of developing cancer, evaluating the  
 CC effectiveness and dosage of a cancer therapy, and determining the  
 CC presence of correlation or non-correlation between activity of at least  
 CC one DNA repair enzyme and a cancer for prevention, early detection and  
 CC prognosis of cancer. ABV72432-43 encode DNA repair/damage preventing  
 CC enzymes, whose level or activity may be determined in the method of the  
 CC invention  
 CC  
 SQ Sequence 1854 BP; 397 A; 546 C; 568 G; 343 T; 0 U; 0 Other;  
 Query Match 10.7%; Score 104; DB 7; Length 1854;  
 Best Local Similarity 54.0%; Pred. No. 2; Se-09;  
 Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;  
 QY 69 GGAGAAGACCCCTTACCGGCTGCTCCGAGTCCCTTCTGACGACGACCCGGGTGA 128  
 DB |||||  
 DB 494 GGACAGCGGGCATATGCTGTGTGCTCAGAGGTCACTGACGACGACCCAGGTTC 553  
 QY 129 GGAAGCCCTCCCTATTACCGCGCTTCTGAGCGCTTCCACCCCTGAAGGCCCTGGC 188  
 DB |||||  
 DB 554 CACTGTGATCACTACTATACCGGATGATGACGAAGTGGCTTACCTGACGACCTGGC 613  
 QY 189 CGCGGCTTCCCTGGAAGAGTCTTAAGGTCTGACGAGGGGCGGCTTACCTGACCGCGGC 248  
 DB |||||  
 DB 614 CAGTGTCTTCTGAGAGAGGTAAATCACTTGGGCTGGCTGGGCTACTATTCCTGTTGG 673  
 QY 249 GGAACACCTCCACCGCCCTGGCCCGAAGCGTGAGAGAGCTTCCCGAGCTTC----- 300  
 DB |||||  
 DB 674 CGGCGGCTGACGAGGAGGCTCGGAAGTGTAGAGAGCTTGAAGGGGCCACATGCCACG 733  
 QY 301 -----GCCGAGCTTCCGGGGGCTTCTGCTGCTGGGCTTACACCGCGGCGCGGT 350  
 DB |||||  
 DB 734 TACAGCAGAGACCCCTGACGAGCTCTGCTGCGTGGGGCGCTTACACAGCTGGGGCAT 793

QY 351 GGCCTCAATCGCCTTTCGGGAGCGGGTGGCGGCGGTGACGGAACGTCCGAGGCTCT 410  
 DB |||||  
 DB 794 TGCCTATCGCCTTTGGCCAGGCAACCGGTGTGTGATGGCAACGTAGCACGGGTGCT 853  
 QY 411 CTCGCGCTCTTTCGGCC--GGGAAAGCCCAAGAGAGAGACTTTTTCGCTGGCCAG 468  
 DB |||||  
 DB 854 GTGCCCTGTCCAGACCATTTGTGCTGATTCACAGACACCTTTTTCGACGACCTCTG 913  
 QY 469 GGCCTCTCTCCCGAAGCGGTGACCG-----GGGTTGGAACAGGCGCTTCAT 518  
 DB |||||  
 DB 914 GGATCTAGCCAGCAGCTGTGAGACCCAGCCCGGACAGAGATTTCACCAAGACCAT 973  
 QY 519 GGAGCTCGGGGCGCACGCTTCCTGCGGAAACGCGCCCGTTGCGGGGCTGCCCCTAGG 578  
 DB |||||  
 DB 974 GGAAGCTAAGGGGACACAGTGTATCCCAACAGCGCCCACTGTGACGACGCTTGGGA 1033  
 QY 579 GGCCTTTCGCGGG 592  
 DB |||||  
 DB 1034 GAGCCTGTGCGGG 1047

RESULT 5  
 AAT89194  
 ID AAT89194 standard; cDNA; 1858 BP.  
 XX  
 AC AAT89194;  
 XX  
 DT 27-MAR-1998 (first entry)  
 XX  
 DE Human mismatch repair MutY cDNA (hMYH gene).  
 XX  
 KM MutY; hMYH gene; mismatch repair; non-polyposis colon cancer;  
 KW xeroderma pigmentosum; gene therapy; diagnosis; human; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PH Key Location/Qualifiers  
 FT CDS 172..1731  
 FT /\*tag= a  
 FT 537  
 FT variation  
 FT /\*tag= b  
 FT /note= "cytosine replaces guanine"  
 FT 852  
 FT /\*tag= c  
 FT /note= "cytosine replaces thymine"  
 FT 1266  
 FT /\*tag= d  
 FT /note= "cytosine replaces guanine"  
 XX  
 PN W09733903-A1.  
 XX  
 PD 18-SEP-1997.  
 XX  
 PF 11-MAR-1996; 96WO-US003239.  
 XX  
 PR 11-MAR-1996; 96WO-US003239.  
 XX  
 PA (SMIT ) SMITTKLINE BEECHAM CORP.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Wei Y;  
 XX  
 DR WPI; 1997-470811/43.  
 DR P-PSDB; AAW31912.  
 XX  
 PT DNA encoding human MutY protein - useful for detecting and treating  
 PT mismatches in DNA especially in non-polyposis colon cancer and xeroderma  
 PT pigmentosum.  
 XX  
 PS Claim 6; Page 37-39; 59pp; English.  
 XX  
 CC This cDNA clone codes for human MutY (see AAW31912), which is homologous



CC to the Escherichia coli MutY protein involved in the pathway that  
CC corrects A/G and A/C mismatches as well as adenines paired with 7,8-  
CC dihydro-8-oxo-deoxyguanine in mutated DNA. It was isolated from a cDNA  
CC library derived from human cerebellum tissue. The hmyh gene contains 15  
CC introns and is 7.1 kb long. Some naturally occurring allelic variants of  
CC the coding sequence have been identified. Polynucleotides at least 70%  
CC identical to the 1811 bp sequence, or complementary to such as sequence,  
CC and comprising at least 30 nucleotides, can be incorporated into a vector  
CC and used for production of recombinant polypeptides in host cells. The  
CC MutY polypeptide, or the polynucleotide encoding it, can be used e.g. to  
CC repair oxidative damage to DNA, to prevent mutations from oxidative  
CC lesions, to treat genetic diseases related to a mutated hmyh gene, e.g.  
CC xeroderma pigmentosum and neoplasia, and to diagnose an abnormal  
CC transformation or a susceptibility to abnormal transformation of cells,  
CC particularly in a non-polyploid colon cancer

XX Sequence 1858 BP; 401 A; 546 C; 567 G; 344 T; 0 U; 0 Other;

Query Match 10.7%; Score 104; DB 2; Length 1858;  
Best Local Similarity 54.0%; Pred. No. 2.5e-09;  
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

Qy 69 GAGAGAGAGACCTTACCGCGCTGCTCGAGAGTCTTTCGAGAGAGACCGCGGTGA 128  
Db 498 GGAAGAGCGGGCAATGCTGTGTGCTGCTGAGAGTATGTCAGAGAGAGAGAGTTGC 557  
Qy 129 GCAGGCGCTCCCTATTATACCGCGCTTTCGAGAGCGCTTTCACCTGAAGCGCTTGC 188  
Db 558 CACTGTATCACTACTATATACCGATGGATGACAGAGGCGCTACACTGAGAGAGCTGGC 617  
Qy 189 CGCGGCTTCCCTGGAAGAGTCTTGAAGTGTGCGAGGGGGGGGCTACTACCGCGGGC 248  
Db 618 CAGTCTTCCCTGAGAGAGTGAATCACTGTGGCTGGCTGCTGCTACTATTTCTGTGG 677  
Qy 249 GGAACACTTCCACCGCGCTGCGAGAGGTGAGAGAGCTTCCCGAGCTTC----- 300  
Db 678 CCGGCGGCTGAGAGAGAGTCCGAGAGTGTAGAGAGTACGAGGGGCCACATGCCACG 737  
Qy 301 -----GCCGAGCTTGGGGGGCTTCTGTGCTCGGGGCTTACACCGCGGGCGGCGGT 350  
Db 738 TACAGAGAGAGCTCTGAGAGAGCTCTGCTGGCTGGGGGGCTACACAGCTGGGGCCAT 797  
Qy 351 GGCTTCATGCTTCCGGGGAGCGGGTGGCGCGCTGAGACGGGAACCTCCGAGAGTCTCT 410  
Db 798 TGCTCTATGCTGCTTGGCCAGGACCGGTGTGTGTGAGTGGCAACGTAGCAGCGGTGCT 857  
Qy 411 CTCGGGCTCTTTCGCC--GGGAAAGCCCCAAGAGAGAGCTTTTGGCCCTGCGCCAG 468  
Db 858 GTGCTGTGTGAGAGCAATGTGTGTGTATCCAGAGACCTTTTGTTCAGGAGCTGTG 917  
Qy 469 GGCCTCTCCCGAGGGGCTGAGCCG-----GGGGTGTGGAACGAGGCTCTCAT 518  
Db 918 GGGGTATAGCCAGAGAGTGTGTGACCCAGCCCGGCAAGATTTCAACAGAGAGCAT 977  
Qy 519 GGAAGTGGGGGAGCGGTCTGCTGCGGAACGGCCCCGTTGCGGGGCTTCCCTTAGG 578  
Db 978 GGAAGTGGGGGAGCGGTGTGTATCCCAAGCGCCCACTGTGAGCCAGTGCCTGTGGA 1037  
Qy 579 GGCCTTGGCGGG 592  
Db 1038 GAGCTGTGCGGG 1051

RESULT 6  
ID AAC78090  
AAC78090 standard; cDNA; 1878 BP.

XX AAC78090;  
XX 08-FEB-2001 (first entry)  
XX Human cancer associated gene sequence SEQ ID NO:484.  
XX DE

KW Human; cancer associated gene; cancer antigen; detection; cancer;  
KW diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;  
KW antidiabetic; antistatic; antirheumatic; antidiabetic; antiviral;  
KW antiinflammatory; antihypertoid; antiallergic; antibacterial; cardiant;  
KW dermatological; neuroprotective; thrombolytic; coagulant; inflammatory;  
KW vasotrophic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
KW allergic reaction; graft versus host disease; organ rejection;  
KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
KW neurological disease; drug screening; ss.

OS Homo sapiens.

PN WO200055350-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US005882.

PR 12-MAR-1999; 99US-0124270P.

PA (HUMA-) HUMAN GENOME SCT INC.

PI Rosen CA, Ruben SM;

DR WPI: 2000-58753/55.

PT P-PSDB; AAB4381.

PS Novel isolated nucleic acids comprising sequences encoding peptides  
useful for treating or diagnosing e.g. cancer.

XX Claim 1; Page 1016-1017; 2352pp; English.

XX AAC78090 to AAC78448 encode the human cancer associated proteins given in  
XX AAB43398 to AAB44239. The proteins can have activities based on the  
XX tissues and cells the genes are expressed in. Example of activities  
XX include: cytostatic; proliferative; vulnerability; immunomodulator;  
XX antidiabetic; antistatic; antirheumatic; antidiabetic; antiviral;  
XX antiinflammatory; antihypertoid; antiallergic; antibacterial; antiviral;  
XX dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
XX vasotrophic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
XX immune disorder; haematopoietic cell disorder; autoimmune disorder;  
XX allergic reaction; graft versus host disease; organ rejection;  
XX haemostatic; thrombolytic; cardiovascular disorder; infection;  
XX neurological disease; drug screening; ss.

XX Sequence 1878 BP; 411 A; 548 C; 566 G; 348 T; 0 U; 5 Other;

Query Match 10.7%; Score 104; DB 3; Length 1878;  
Best Local Similarity 54.0%; Pred. No. 2.5e-09;  
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

Qy 69 GAGAGAGAGACCTTACCGCGCTGCTCGAGAGTCTTTCGAGAGAGACCGCGGTGA 128  
Db 509 GGAAGAGCGGGCAATGCTGTGTGCTGCTGAGAGTATGTCAGAGAGAGAGAGTTGC 568  
Qy 129 GCAGGCGCTCCCTATTATACCGCGCTTTCGAGAGCGCTTTCACCTGAAGCGCTTGC 188  
Db 569 CACTGTATCACTACTATATACCGATGGATGACAGAGTGTGCTCATCTGAGAGAGCTTGC 628  
Qy 189 CGCGGCTTCCCTGGAAGAGTCTTGAAGTGTGCGAGGGGGGGGCTACTACCGCGGGC 248  
Db 629 CAGTCTTCCCTGAGAGAGTGAATCACTGTGGCTGGCGCTGCTGCTACTATTTCTGTGG 688

```
QY 249 GGAACACTTCCACCGGCTGGCCGGAAGCGTGGAGAGCTTCCCGAGCTT----- 300
Db 689 CCGGCGGCTGCAGAGAGAGAGCTGGAAGGTGTAAGAGAGCTAGGGGCGACATGCGCAG 748
QY 301 -----GCCGAGCTTCGAGGAGCTTCTGTGCTCGGCTTACACCGCGCGCGGT 350
Db 749 TACAGCAGAGAGACCTGACAGAGCTCTGCGCTGGCGTGGGCGCTACACAGCTGGGCGCAT 808
QY 351 GGCCTTCATTCGCTTGGGAGAGCGGCTGGCGCGGTGAGACGGGAAGTCCGAGGGTCTT 410
Db 809 TGCCTTATTCGCTTGTGGCCAGGCAACCGGTGTGTGATGTGCACAGTACAGGCTGCT 868
QY 411 CTCGCCCTCTTTCGCCC--GGGAAAGCCCCAAGAGAGAGCTTTCGCGCTCGCCAG 468
Db 869 GTGCCCTGTCCAGGCAATTTGGTCTGATCCAGCAGACCTTTGTTCCAGAGCTCTG 928
QY 469 GGCCTCTCCCGAGGCGGTGACCCG-----GGGTTGTGAAACGAGCCCTCAT 518
Db 929 GGGTTAGCCCAAGCAGCTGTGTGACCCAGCCGCGCAGAGATTCAACCAAGCAGCAT 988
QY 519 GAGCTTCGGGGCCACGCTCTGCTGCTGCCGAAACGAGCCCGTTGGGGGCTGCCCCCTAG 578
Db 989 GAGGCTAAGGGGCCACAGTGTGTATCCCAAGCGCCCATGTGCAGCAGTGCCTGTGA 1048
QY 579 GGCCTTCGCGGG 592
Db 1049 GAGCTGTGCGGG 1062
```

```
RESULT 7
AAV35701
ID AAV35701 standard; cDNA; 1856 BP.
```

```
AC AAV35701;
```

```
XX 19-AUG-1998 (first entry)
```

```
XX cDNA encoding human MYH.
```

```
XX Human; MYH; hMYH; diagnosis; cancer; ss.
```

```
XX Homo sapiens.
```

```
XX Key Location/Qualifiers
```

```
XX FT 172..1776
```

```
XX FT CDS /tag= a "MYH"
```

```
XX FT /product= "MYH"
```

```
XX JPI0057076-A.
```

```
XX 03-MAR-1998.
```

```
XX 11-MAR-1997; 97JP-00099540.
```

```
XX 11-MAR-1996; 96US-0013132P.
```

```
XX (HUMA-) HUMAN GENOME SCI INC.
```

```
XX WPI; 1998-210407/19.
```

```
XX P-PSDB; AAW60243.
```

```
XX New isolated polypeptide - useful for, e.g. diagnosis of cancer.
```

```
XX Claim 6; Page 17-19; 23pp; Japanese.
```

```
XX The present sequence encodes human MYH (hMYH). The MYH polypeptide can be
```

```
XX used for diagnosing cancer, where the decreasing levels of the
```

```
XX polypeptide are measured in a sample from a patient
```

```
XX Sequence 1856 BP; 400 A; 545 C; 568 G; 343 T; 0 U; 0 Other;
```

```
Query Match 10.5%; Score 102.6; DB 2; Length 1856;
```

```
Best Local Similarity 54.4%; Pred. No. 4.4e-09;
```

```
Matches 291; Conservative 0; Mismatches 214; Indels 30; Gaps 3;
QY 88 GTCTGTCTCCGAGGCTCTTCTGACAGACCCGGGTGAGAGGCTCCCTATTAC 147
Db 514 GTGTGGTCTCAGAGGTATGATCTGACAGACCCAGGTTCACATGTATCACTATAT 573
QY 148 GCGCGCTTTTGGAGCGCTTTCCACCCGTAAGGCCCTGGCGGCTTCCTGGAGAG 207
Db 574 ACCGATGATGCAAGATGAGCTTACACTGACAGACCTGGGCAAGTCTTCTGGAGAG 633
QY 208 GTCTTAAAGGTCTGACAGGAGGCTTACACCGGCGGAGAACCTTCACCGCTG 267
Db 634 GTGATCAACTCTGGGCTGGCTGGCTGCTATTTCTGTGGCCGGCTGCAAGAGGA 693
QY 268 GCCGAGCGTGGAGAGAGCTTCCCGAGCTT-----GCCGAGCTT 309
Db 694 GCTCGAAGAGTGTAGAGAGCTAGGGGGCCACATGCCACGTAGACAGAGACCTTGCAG 753
QY 310 CCGGGGCTTCTGCTGCTGGGCTTACACCGGCGGCGGCTGCTCATGCGCTTGGG 369
Db 754 CAGCTTCTGCTGGGCTGGGCGCTACACAGCTGGGCGCATTTGCTTATGCTTGGC 813
QY 370 GAGCGGTGCGCGGTGAGACGGGAACCTCCGAGGCTCTTCCGCTTTCGCC-- 427
Db 814 CAGGCAACCGGTGTGTGATGGCAAGTACAGGGGTCTGTGCCGTTCGAGCATT 873
QY 428 GGGAAAGCCCCAAGAGAGAGCTTTTCCGCTTCGCCAGGCTCTCTCCGAGGCG 487
Db 874 GGTGTGATCCACAGACGACCTGTGTTCCAGCAGCTGTGGGTCTAGCCACAGCTG 933
QY 488 TGG-----ACCGGGGTGTGGAACCAAGGCGCTCATGAGCTCGGGGCCAGGTC 537
Db 934 GTGACCCAGCCCGGCCAGAGATTTCACCAAGCAGCAGTGAAGCTTAGGGGCCACAGTG 993
QY 538 TGCCTGCGAAACGAGCCCGCTTGGCGGCGCTGCCCTTAGGGGCTTCTGCGGG 592
Db 994 TGTACCCACAGCGGCCCATGTGACAGCCAGTGGCCCTGTGAGAGACCTGTGCCGG 1048
```

```
RESULT 8
```

```
ADBE36257
ID ADBE36257 standard; cDNA; 1608 BP.
```

```
AC ADBE36257;
```

```
XX 29-JAN-2004 (first entry)
```

```
XX Human MYH (hMYH) A494G mutant cDNA.
```

```
XX Gene; ss; mutant; human; screening method; hMYH; base excision repair;
```

```
XX BBR; APC; familial adenomatous polyposis; FAP;
```

```
XX multiple colorectal adenoma; carcinoma; bowel cancer.
```

```
XX Homo sapiens.
```

```
XX Key Location/Qualifiers
```

```
XX FT 1..1608
```

```
XX FT /tag= a
```

```
XX FT /product= "Human MYH mutant protein (Y165C)"
```

```
XX FT allele
```

```
XX FT /tag= b
```

```
XX FT /note= "Base G in mutant replaces wild type base A"
```

```
XX WO2003014390-A2.
```

```
XX 20-FEB-2003.
```

```
XX 02-AUG-2002; 2002WO-GB003591.
```

```
XX 03-AUG-2001; 2001GB-00018995.
```

```
XX (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
```

PI Sampson JR, Cheadle JP;  
 XX WPI: 2003-256601/25.  
 DR P-PSDB; ADE36260.  
 XX  
 PT Screening, diagnostic and therapeutic methods in individuals with  
 PT predisposition towards having a cancer, such as colon cancer, using base  
 PT excision repair pathway or hMYH genes.  
 XX  
 PS Claim 8; SEQ ID NO 1; 66pp; English.  
 XX  
 CC This invention relates to a novel screening method for identifying an  
 CC individual having a predisposition towards a cancer. Specifically, it  
 CC refers to obtaining a test sample, preferably comprising the hMYH gene  
 CC that occurs in the base excision repair (BER) pathway, and comparing this  
 CC nucleic acid molecule to the corresponding region of the wild type  
 CC sequence. This BER pathway gene, hMYH, acts to protect against G:C to T:A  
 CC transverse mutations in a cancer marker gene such as APC that is seen in  
 CC familial adenomatous polyposis (FAP). As such, mutations identified in  
 CC hMYH are associated with the onset multiple colorectal adenomas and  
 CC carcinoma. The present invention describes a screening method for  
 CC individuals that works to identify differences comprising any one of  
 CC G382D, Y165C, E466X or Y90X variations in hMYH, this signifies a cancer  
 CC predisposition, particularly for bowel cancer. This polymucleotide  
 CC sequence is the mutant A494G hMYH cDNA that encodes a Y165C mutation in  
 CC the encoded polypeptide of the invention.  
 CC  
 XX Sequence 1608 BP; 336 A; 477 C; 504 G; 290 T; 0 U; 1 Other;  
 SQ  
 Query Match 10.5%; Score 102; DB 9; Length 1608;  
 Best Local Similarity 53.6%; Pred. No. 5.7e-09;  
 Matches 297; Conservative 1; Mismatches 226; Indels 30; Gaps 3;  
 QY 69 GAGAGAGACCCCTTACCGCGTCTGCTCCGAGGCTCTTTCAGCAGACCCGGGTGA 128  
 DB 327 GAGACGGGGGATATGCTGTGTGTCTCAGAGGTCTGACAGACCCAGGTTC 386  
 QY 129 GCAGGCTTCCCTTATACCGCGCTTTTTCAGCGCTTTCCACCTGAAAGCCCTGGC 188  
 DB 387 CACTGTGATCACTACTATACCGAGTGTATGACAGAGGCGCTTACCTGACGAGACTTGGC 446  
 QY 189 CGCGGCTTCCCTGAGAGAGTCTTATAGGCTGTGCGAGGGGGGCGCTACTCCCGCGGC 248  
 DB 447 CAGTCTTCTCTGAGAGAGGTGATCACTTGGGCTTGGCTGTATTTCTCTGCG 506  
 QY 249 GAAACACCTCCACCGCTGCGCCGAAAGGTGAGAGCTTCCCGAGCTTC----- 300  
 DB 507 CCGCGGCTGACAGAGGAGCTCGGAGGTGTAGAGAGAGGAGGCGCACATGCCACG 566  
 QY 301 -----GCCGAGCTTGGGGGCTTCTGTGTCTCGGCGCTTACACCGCGCGCGCT 350  
 DB 567 TACAGCAGAGACCTCTGACGAGCTCTGCTGCGGCGGCGCTACACAGCTGGGGCCAT 626  
 QY 351 GGCTTCATGCGCTTGGGGGAGCGGGTGGCGGTGACGAGGAACTCCGAGGGTCTCT 410  
 DB 627 TGCTCTATGCTCTTGGCCAGGCAACCGGTGTGTGATG3CAACGAGACGGGTGCT 686  
 QY 411 CTCGCGCTCTTTCGCC--GGGAAAGCCCAAGAGAGAGCTTTTGGCCCTGCGCCAG 468  
 DB 687 GTGCGGTGTGCGAGGCAATGTGTGATCCAGACGACCTTGTGTTCCGACGAGCTTG 746  
 QY 469 GGCCTCTCTCCGAGGGCGTGAACCG--GGGTGTGGAACAGAGCCCTCAT 518  
 DB 747 GGGTCTAGCCCAAGAGCTGTGTGAGACCCAGCCGCGCAAGAAATTTCAACAGAGCCAT 806  
 QY 519 GGAGCTGGGGGCAACGCTGTGCTGCGAAACGGCCCGTGTGGGGGCTGCCCCCTTAG 578  
 DB 807 GGAAGTAAAGGACAGTGTGTATCCCAAGAGCCCACTGTGACAGCCAGTGCCTTGA 866  
 QY 579 GGCCTTTCGCGGG 592  
 DB 867 GAGCTGTGCGGG 880

RESULT 9  
 ID AAF21610 standard; DNA; 349980 BP.  
 XX  
 AC AAF21610;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 KW ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN MO20006791-A1.  
 XX  
 PD 09-NOV-2000.  
 XX  
 XX 08-MAR-2000; 2000WO-US005928.  
 XX  
 PF 30-APR-1999; 99US-0132068P.  
 PR 08-OCT-1999; 99WO-US023573.  
 PR 28-FEB-2000; 2000GB-00004695.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Scarlato V,  
 PI Maignani V, Galeotti C, Mora M, Ratti G, Scarfelli M, Scarlato V,  
 PI Rappuoli R, Frazer CM, Grandi G;  
 XX  
 DR WPI: 2000-647603/62.  
 XX  
 PT Neisseria meningitidis B full length genome sequence and open reading  
 PT frames are used to detect, treat and prevent Neisserial infections.  
 XX  
 PS Claim 7; Appendix A; 692pp; English.  
 XX  
 CC The present invention describes the full length genome of Neisseria  
 CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613  
 CC represent fragments of the NMB genomic sequence, as the sequences which  
 CC long to go in a record on its own it was split into 8 sequences which  
 CC overlap each other at the beginning and end of each sequence by 49980 bp  
 CC (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of  
 CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of  
 CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins  
 CC given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR  
 CC primers which are used in the exemplification of the present invention.  
 CC The NMB genome and fragments from it have antibacterial activity, and can  
 CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins  
 CC and/or antibodies which binds to the proteins can be used in compositions  
 CC for treating or preventing infection due to Neisserial bacteria or as a  
 CC diagnostic reagent for detecting the presence of Neisserial bacteria or  
 CC of antibodies raised for Neisserial bacteria. Computers, computer memory,  
 CC computer storage medium or computer databases can be used in a search to  
 CC identify open reading frames (ORFs) or coding sequences within the NMB  
 CC genome. The DNA sequences provide further opportunities to find antigenic  
 CC or immunogenic proteins which are more effective in vaccines than the  
 CC outer membrane proteins currently used  
 CC  
 XX Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 U; 0 Other;  
 SQ  
 Query Match 9.8%; Score 95.6; DB 3; Length 349980;  
 Best Local Similarity 51.6%; Pred. No. 5.8e-08;  
 Matches 255; Conservative 0; Mismatches 224; Indels 15; Gaps 1;  
 QY 17 AAGCCTCTCTGCTGTGATCCGGAAGAAACGCCCGCTCTGCGGGGAGAGG 76  
 DB 225615 AAGGCTCATCTCGCTGGCAAAACACACCGGTCTCCACCACTCTGCGAGGTCAAAA 225674

QY 77 ACCCTTACCGGCTCTGCTCTCCGAGTCTCTTCTGACAGACCCGGGTGAGAGGCCC 136  
 Db 225675 ACCCTTATTTGGCTCTGGCTTTCCGAAATATGCTCCAGCAAAACGAAATGCGCACCGTGT 225734  
 QY 137 TCCCTTATTTACCGGCTCTTCTGAGCGCTTTCCACCTTAAAGCCCTTGGCCGGGCTT 196  
 Db 225735 TGGACTACTATCCGGCTTCTTGAATAATTCGACCGTTTGAACGCTTGGCCGGCGCG 225794  
 QY 197 CCTTGAAGAGTCTCTTGAAGGCTCTGACAGGGGCGGGCTACTACGGCGGGCGGAACCC 256  
 Db 225795 CCGAAGACCAAGTGTGTGTGTGTGGCGGCTTGGGCTATTACAGCCGGCGCGCAAC 225854  
 QY 257 TCCACCGCTGGCGCCGCAACGCTGAGAGAC-----TTCCCGGAGCTTGC 301  
 Db 225855 TGCACAAAGCCCGCAACAAAGTGTGAGCAATTCGGCGGCACTTTCGCTGAGAGCGCA 225914  
 QY 302 CCGAGCTTGGGGGCTTCTCTGTCTCTGGGCTTACACCGCGCGCGCGGCTCTCCATCG 361  
 Db 225915 AAGACTTGGAAACCTCTCTGGCGGTAGCAAGACCGCGCGCAATTTGGCGGCTTCT 225974  
 QY 362 CCTTGGAGGAGGGGGTGGCGGGGTGACGGGAACTCCGAGAGGCTCTCTCCGCTCT 421  
 Db 225975 CCTTCAACCGCGCGCAACATTTTGAAGCACTCAAGCGCTACTCTGCGCGGTGT 226034  
 QY 422 TCGCCCGGGAAGCCCAAG 481  
 Db 226035 TCGCCCGGAGCGCAATCCGAGAGCAAAATTTGAAATTCGCTTGTGACACTTGGCG 226094  
 QY 482 AGGGCTGTGACCGC 495  
 Db 226095 AAGCCTGTGCTCG 226108

RESULT 10  
 AAA81490\_13  
 Continuation (14 of 15) of AAA81490 from base 1300001 (N. meningitidis B full length gen  
 WP Sequence split into 15 fragments LOCUS AAA81490 Accession Aaa81490

Fragment Name	Begin	End
WP AAA81490_00	1	110000
WP AAA81490_01	100001	210000
WP AAA81490_02	200001	310000
WP AAA81490_03	300001	410000
WP AAA81490_04	400001	510000
WP AAA81490_05	500001	610000
WP AAA81490_06	600001	710000
WP AAA81490_07	700001	810000
WP AAA81490_08	800001	910000
WP AAA81490_09	900001	1010000
WP AAA81490_10	1000001	1110000
WP AAA81490_11	1100001	1210000
WP AAA81490_12	1200001	1310000
WP AAA81490_13	1300001	1410000
WP AAA81490_14	1400001	1437668

Query Match 9.8%; Score 95.4; DB 3; Length 110000;  
 Best Local Similarity 51.7%; Pred. No. 6.6e-08;  
 Matches 253; Conservative 0; Mismatches 221; Indels 15; Gaps 1;

QY 22 CTCTCTGCTGTGTACCGGAAACGCCGCCCTCTCTCTGCGGGGGAAGAGACCT 81  
 Db 96008 CTCTCTGCTGTGTGCAAAACAAACAGGTGCGCACACCTCTCTGAGGTCAAAAACCT 96067  
 QY 82 TACCGGCTCTGTCTCCGAGTCTTCTGACAGACCCGGGTGAGAGAGCCCTCC 141  
 Db 96068 TATGCGCTGTGCTTCCGAATATCTCTCAAGCAACGAAAGTGCACCGTGTGAC 96127  
 QY 142 TATTAACGCGCTTCTGAGAGCGCTTCCACCTGAAGGCCCTGAGCGGCTTCCCTG 201  
 Db 96128 TACTATCCGCGCTTCTTGAATAATTCGACCGCTTCAAGAGCTTGCCTGCGCGCGCA 96187  
 QY 202 GAAGAGTCTTAGAGTCTGACAGGGGCGGCTACTACCGCGGGCGGAACACTTCCAC 261  
 Db 96188 GACGAGTGTGTGTGTGTGTGGCGGCTTGGGCTATTACAGCGCGCGCAACCTGCAC 96247

QY 262 CGCTGCGCCGAAACCTGTGAGAGC-----TTCCCGGAGCTTGGCGAG 306  
 Db 96248 AAAGCCCGCAACAAAGTGTGACAGCAATTCGCGGCACTTTCGCTGAGCGCAAGAC 96307  
 QY 307 CTTCGGGGGCTTCTGTGTCTGAGGCTTACACCGCGCGCGGTGAGCTTCATGCGCTTC 366  
 Db 96308 TTGAAACCTCTTGGCGCGTATGACAGAGACACCGCGCGCTCATTTGCGCTTCTCCTTC 96367  
 QY 367 GGGAGCGGTGGCGGCGGTGACCGGAACTCCGAGGGTCTCTTCCGCTTTCGCC 426  
 Db 96368 AACCGCGCAACATTTTGAACGCGCAACGTCAACGCTACTTGGCGGCTTTCGCC 96427  
 QY 427 CCGAAGAGCCCAAG 486  
 Db 96428 CGCAGCGCAATCCGAGAGCAAAATTTGAAATTCGCTTGTGACACTTTCGCAAGC 96487  
 QY 487 GTGAGCCCG 495  
 Db 96488 CTGCTGCCG 96496

RESULT 11  
 ABK74292  
 ID ABK74292 standard; DNA; 528 BP.  
 XX AC ABK74292;  
 XX 13-AUG-2002 (first entry)  
 DT Bacillus licheniformis genomic sequence tag (GST) #1583.  
 DE Bacillus licheniformis genomic sequence tag (GST) #1583.  
 KW Differential gene expression; genomic sequenced tag; GST;  
 KW altered culture condition; environmental stress;  
 KW physiological provocation; ds.  
 OS Bacillus licheniformis.  
 XX  
 PN WO200229113-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 05-OCT-2001; 2001WO-US031437.  
 XX  
 PR 06-OCT-2000; 2000US-00680598.  
 XX  
 PR 27-MAR-2001; 2001US-0279526P.  
 XX  
 PA (NOVO) NOVOZYMES BIOTECH INC.  
 XX  
 PI (NOVO) NOVOZYMES AS.  
 XX  
 PI Berke R. Clausen IG;  
 XX  
 DR WPI; 2002-416684/44.  
 XX  
 PT Monitoring differential expression of several genes in first Bacillus  
 PT cell relative to expression of same genes in one or more second Bacillus  
 PT cells, by using substrate containing Bacillus genomic sequenced tag  
 array.  
 XX  
 PS Claim 4; SEQ ID NO 1583; 200pp; English.  
 XX  
 PS The invention describes a method of monitoring differential expression of  
 CC genes in a first Bacillus cell relative to expression of the genes in  
 CC other Bacillus cells, comprising hybridizing labelled nucleic acid probes  
 CC isolated from Bacillus cells to a substrate containing array of Bacillus  
 CC genomic sequenced tags (GST), examining the array, and determining  
 CC relative gene expression by an observed hybridisation reporter signal of  
 CC a spot in the array. The method is useful for measuring the expression of  
 CC genes in a first Bacillus cell relative to expression of the same genes  
 CC in one or more second Bacillus cells. The method is useful for monitoring  
 CC global expression of several genes from a Bacillus cell, discovering new  
 CC genes, identifying possible functions of unknown open reading frames and  
 CC monitoring gene copy number variation and stability. Monitoring changes

CC in expression of genes may be used to provide a representation of the way  
CC in which *Bacillus* cells adapt to changes in culture conditions, exten  
CC environmental stress or other physiological provocation. Extensive follow  
CC -up characterisation is unnecessary, when one spot on an array equals one  
CC gene or one open reading frame, since sequence information is available.  
CC This sequence represents a genomic sequence tag (GST) used in the method  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPRO at ftp.wipro.int/pub/published\_pcc\_sequences

XX Sequence 528 BP; 160 A; 111 C; 140 G; 116 T; 0 U; 1 Other;

Query Match 9.5%; Score 92.8; DB 6; Length 528;

Best Local Similarity 53.4%; Pred. No. 2.3e-07;

Matches 227; Conservative 0; Mismatches 183; Indels 15; Gaps 1;

QY 17 AAGCCCTCTGCGCTGTACCGGGAAAACGCCGCCCTCCCTGCGGGGGGAGAG 76  
DB 43 AAGATTATTTCTTGATGAGCAAGAAAACGGGATCTGCTTGGCGCTCAGACAGG 102  
QY 77 ACCCTTACCGGCTCTGCTCCGAGCTCTTTCAGACAGACCCGGGTGAGCAGGCC 136  
DB 103 ACCCTTATAGGTGTGGGTGTGCGAAGTATGCTTCAACAGACGAGGTGACACGGTGA 162  
QY 137 TCCCTATTACCGCGCTTCTGAGCGCTTCCACCGCTGAAGGCCCTGCGCGCGCTT 196  
DB 163 TTCTTACTTCAACAACCTTATCGAAAAGTTTCGACCGGTGAGCGCTCGCTGAAGCG 222  
QY 197 CCTTGAAGAGGCTCTTGAAGGTCTGCGAGGGCGGCGCTACTACCGCGCGGCGAACC 256  
DB 223 ACGAAGAAAAGTGTCTTAAAGCTTGGGAAGACTGGATATCTCAAGGGTTGGAACT 282  
QY 257 TCACACCG-----CTGCCCGAAGCTGAGAGCTTCCCGGAGCTTCC 301  
DB 283 TGCAGAGCGCTGTACAGGAAGTTCAAGACATACGAGGTGTGCTCCCTTCACAAAG 342  
QY 302 CCGAGCTTCGGGGGCTTCTGCTCGGGCCCTTACACCGCGCGCGGTGGCTTCATCG 361  
DB 343 AGGAATTGCGCAGCTTAAAGGCGTGGCCCTTATACAAAGCGCGGTGCTGAGCATCG 402  
QY 362 CTTTCGGGAGCGGCTGCGGCGGTGAGACGGGAACGTCGGAAGGCTCTTCCGCTCT 421  
DB 403 CCTTCAATCAGCGGCTCCCGCGGTAGACGGAATGTATGCGGCTCATGTACGAGATT 462  
QY 422 TCGCC 426  
DB 463 TTTC 467

RESULT 12  
ABZ40024  
ID ABZ40024 standard; DNA; 1047 BP.

XX ABZ40024;

XX 07-MAR-2003 (first entry)

XX N. gonorrhoeae nucleotide sequence SEQ ID 4637.

XX Antibacterial; infection; vaccine; gene therapy; gene; de.

XX *Neisseria gonorrhoeae*.

XX MO200279243-A2.

XX 10-OCT-2002.

XX 12-FEB-2002; 2002WO-IB002069.

XX 12-FEB-2001; 2001GB-00003424.

XX (CHIR-) CHIRON SPA.

XX PA

PI Fontana MR, Pizza M, Maignani V, Monaci E;  
XX WPI; 2003-058415/05.  
DR P-PSDB; ABP79054.

XX New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a  
XX medicament for treating or preventing *N. gonorrhoeae* infection.

XX Disclosure; Page 528; 815pp; English.

XX The present invention relates to proteins from *Neisseria gonorrhoeae*.

XX Also disclosed are the nucleic acid molecules encoding the proteins and  
XX antibodies that specifically bind to the proteins. The composition  
XX comprising the protein, nucleic acid or antibody is useful for the  
XX manufacture of a medicament for treating or preventing *N. gonorrhoeae*  
XX infection, this may be in the form of a vaccine or gene therapy.

XX Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
XX molecules of the invention

XX Sequence 1047 BP; 259 A; 306 C; 262 G; 220 T; 0 U; 0 Other;

XX Query Match 9.1%; Score 89.2; DB 7; Length 1047;

XX Best Local Similarity 50.8%; Pred. No. 9.5e-07;

XX Matches 251; Conservative 0; Mismatches 228; Indels 15; Gaps 1;

QY 17 AAGCCCTCTGCGCTGTACCGGGAAAACGCCGCCCTCCCTGCGGGGGGAGAG 76  
DB 35 AAGGCTCATCGCTGCGAAAACACAGGTGTGCGACACCTCTTGGCAGGTCAAA 94  
QY 77 ACCCTTACCGGCTCTGCTCCGAGCTCTTTCAGACAGACCCGGGTGAGCAGGCC 136  
DB 95 ACCCTTATGCGCTCTGCTTCCGAATCATGCTCAGCAACGAGGTTCGCCCGGT 154  
QY 137 TCCCTATTACCGCGCTTCTGAGCGCTTCCACCTGAAGCCCTGCGCGCGCTT 196  
DB 155 TGACTACTATCCGCTTCTTGGAAAATTCGACCGCTTCAAGACGCTTCCCGCGCG 214  
QY 197 CCTTGAAGAGTCTCTTGAAGGTCTGCGAGGGCGGCGGTACTACCGGGGGCGGAACC 256  
DB 215 CGAAGAGAGAGTGTGTGTGTGGGGGCTTGGGTATTTACGCGCGCGCGCAATC 274  
QY 257 TCAC-----CGCTGCGCGAAGCTGAGAGCTTCCCGGAGCTTCC 301  
DB 275 TGCACAAAGCGCGCACAAATCTGTGACCAATTCGGGGTACGTTTCAATCGAGCCCA 334  
QY 302 CCGAGCTTCGGGGGCTTCTGCTCGGGCCCTTACACCGCGCGCGGTGCTTCATCG 361  
DB 335 AAGACTTGAAGACGCTGCGGGCGTGAAGCAGACACGCGCGCAATTTCTGCTTTG 394  
QY 362 CTTTCGGGAGCGGCTGCGGCGGTGAGACGGGAACGTCGGAAGGCTCTTCCCGCTCT 421  
DB 395 CTTTCAACCGACGAGAAACATCTTGAACGCGCAACGCGTCTTCTGCGGTGTT 454  
QY 422 TCGCCCGGAAAGCCCAAGAGAGATTTCGCTCGCCCTGCGGAGGCTCTTCCCGCG 481  
DB 455 TTGCCCAAGACGGCAATCCGACATTAATAAATTGAAAACTCGCTCTGACACTTGGCG 514  
QY 482 AGGCGGTGACCCG 495  
DB 515 AAGCGCTGATGCGG 528

RESULT 13  
AAS63242  
ID AAS63242 standard; CDNA; 2293 BP.

XX AAS63242;

XX 29-JAN-2002 (first entry)

XX *Escherichia coli* (strain K-12) A/G-specific adenine glycosylase MutY DNA.

XX DNA mutation-binding protein; nuclease; DNA mismatch; cancer; PCR primer;

XX KM

KM DNA damage; human xeroderma pigmentosum complementation group; XPE; XPA;  
KM XPC; XPE; BECC4; human Muts homologue 2; hMSH2; Muts; Nuc; MutY; Fpg; ss;  
KM Fpgy-DNA glycosylase; uracil DNA glycosylase; ung; TDG; xtha gene; Uvr A;  
KM A/G-specific adenine glycosylase; synthetic T4 endonuclease V; T4 endo V;  
KM thymine DNA-glycosylase; Uvr B; Uvr C; nfh gene; nfo gene; exonuclease;  
endonuclease.  
OS  
XX Escherichia coli.  
XX  
XX MO200173079-A2.  
XX  
XX 04-OCT-2001.  
XX  
XX 26-MAR-2001; 2001WO-US009700.  
XX  
XX 28-MAR-2000; 2000US-0192764P.  
XX PR 29-AUG-2000; 2000US-00650855.  
XX  
XX (REGC ) UNITV CALIFORNIA.  
XX  
XX Mc Cutchen-Malone SL;  
XX  
XX WPI: 2001-656920/75.  
XX DR P-PSDB; AAN69755.  
XX  
XX Recombinant chimeric protein, useful for detecting and quantifying DNA  
PT mutations, e.g. in disease diagnosis, comprises mutation-binding protein  
PT and nuclease.  
XX  
XX Claim 58; Page 99-100; 128pp; English.  
XX  
XX Sequences AAS63224-AAS63305 represent cDNA encoding and PCR primers used  
CC to amplify cDNA encoding proteins which can be used in the synthesis of  
CC chimeric proteins comprising a DNA mutation-binding protein, a linker and  
CC a nuclease, by recombinant technology. The chimeric proteins are useful  
CC for detection, quantification and mapping of DNA sequence variations  
CC including mutations, for example, caused by damage and mismatches. The  
CC proteins are able to bind to the site of the DNA mutation and cut it out  
CC of the molecule. This is useful for early diagnosis of cancer and other  
CC diseases. The proteins used in the invention include human XPF (or  
CC ERCC4), human xeroderma pigmentosum complementation groups A, C and E  
CC (XPA, XPC and XPE), human Muts homologue 2 (hMSH2), Serratia marcescens  
CC nuclease (Nuc), Thermus thermophilus Muts, Escherichia coli Fpgy-DNA  
CC glycosylase (Fpg), uracil DNA glycosylase (ung), A/G-specific adenine  
CC glycosylase (MutY), synthetic T4 endonuclease V (T4 endo V), thymine DNA-  
CC glycosylase (TDS), E. coli Uvr A, B and C, and E. coli endonucleases and  
CC exonucleases  
XX  
SQ Sequence 2293 BP; 533 A; 620 C; 593 G; 547 T; 0 U; 0 Other;  
Query Match 8.9%; Score 87.2; DB 4; Length 2293;  
Best Local Similarity 54.1%; Pred. No. 2e-06;  
Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;  
QY 53 CCTTCCCTGGCGGGGGAAGAGACCTTACCGCTCTCTCCAGAGCTTCTTGC 112  
DB 1161 CTCTGCTCTGGCAATTGACAAGACGCTTACAAAGTAGGTCTCAGAAGTATGTC 1220  
QY 113 AGCAGACCGGGGGGAGCAGCGCTTCTTATCCGCGCTTCTTGAAGCGCTTCCCA 172  
DB 1221 AACAACTAGAGGTGGACCGTTATCCCTATTGGAACGCTTATGCGCGCTTCCGA 1280  
QY 173 CCTGAAGCGCTTGGCGCGGCTTCCGGAAGAGGTCTTAAAGGTCTGAGAGGGGGCGG 232  
DB 1281 CGGTGACCGATCTCGCCATATGCGCGCTCGAGGAAGTCTTCACTTGTGAGCCGGCTTG 1340  
QY 233 GTTACTACGGGGGCGGAACCTTCAACCGCTTGGCCGGAAGCGTGGAG----- 282  
DB 1341 GCTATTATGCGCGCGCGCAATCTGCAATTAAGCGGCAACAAGTGGCGACTTACAG 1400  
QY 283 -----GAGTTTCCCGGAGCTTCCGCGAGCTTCCGGGGGCTTCGAGTCTCGGGCTTACA 337  
DB 1401 GCGGTAAATTCGCGAACCCTTTAGAGAGATTGACGACCTGCGGGGCGTGGGCTTCCA 1460

QY 338 CCGGCGGCGGCGGCGCTTCATTCGCTTGGGAGACGGGTGCGCGGTGACGGGAACG 397  
DB 1461 CCGCAGCGCGGATTTCTTCTGCTTCTGCTTGGTAAACACTTTCGATTCGACGGTAACG 1520  
QY 398 TCCGAGAGGTCTCTCCGCGCTTTCGC 425  
DB 1521 TCMAACGCTGCTGGCGCGCTGCTATGC 1548  
RESULT 14  
ABZ79956  
ID ABZ79956 standard; DNA; 912 BP.  
XX  
XX AC ABZ79956;  
XX  
XX 19-MAY-2003 (first entry)  
XX  
XX DE Mycobacterium tuberculosis mutY ORF sequence SEQ ID NO:31.  
XX  
XX KM Mycobacterium tuberculosis; mutY2; alkA; cgt; Rv3908; mutY; Rv3909;  
KM detection; multidrug resistance; multiple drug resistance; MDR;  
XX infection; gene; ds.  
XX  
XX OS Mycobacterium tuberculosis.  
XX  
XX WO2003016562-A2.  
XX  
XX 27-FEB-2003.  
XX  
XX 14-AUG-2002; 2002WO-EP009679.  
XX  
XX PR 14-AUG-2001; 2001US-0311824P.  
XX PR 21-AUG-2001; 2001US-0313523P.  
XX  
XX (INSP ) INST PASTEUR.  
XX  
XX PI Gicquel B;  
XX  
XX WPI: 2003-256711/25.  
XX  
XX PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate  
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate  
PT by detecting an alteration in the DNA repair system of the isolate.  
XX  
XX Claim 45; Fig 6B; 83pp; English.  
XX  
XX The present invention describes a method for predicting the epidemic  
CC character of a Mycobacterium tuberculosis isolate and/or a selective  
CC advantage to be maintained in the host and/or the acquisition of multiple  
CC drug resistance (MDR) by the isolate comprising detecting an alteration  
CC in the DNA repair system of the isolate. Also described: (1) detecting a  
CC polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (2) a  
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a M. tuberculosis mutY open reading frame  
CC (ORF) sequence, which is used in the exemplification of the present  
CC invention  
XX  
SQ Sequence 912 BP; 124 A; 311 C; 326 G; 151 T; 0 U; 0 Other;  
Query Match 8.9%; Score 86.6; DB 7; Length 912;  
Best Local Similarity 50.2%; Pred. No. 2.7e-06;  
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;  
QY 22 CTCTCTGCTGTATCCGGGAAAGACCGCGCGCTTCCCTCGGCGGGG-----GAGAAAGAC 78  
DB 61 CTCTCGCTGTATCAGCATTCGACCGGAGTCTGCTCGGAGAGCCCGGTGTACG 120

```
QY 79 CTTACCGCGCTGCTCCGAGGTCCTTGTGACAGACCCCGGAGGAGCCCTC 138
DB 121 CCGTGGCGATCTGTGTACGAGTTCACTGTCAGCAGACGCCCGCCCGGGGTGCTG 180
QY 139 CCTATTACCGCGCTTCTGTGAGCGCTTTCACCTGAAAGGCCCTGGCGGCTTCC 198
DB 181 GCGATCTGGCCGGAATGGGTGGCGGTGGCCACCGCGTGGCCACCGCCACGCGCAGC 240
QY 199 CTGGAAGAGTCTTAGGCTTGTGCGAGGGGCGGCTCTAACCGCGGCGGAAACCTC 258
DB 241 ACCCGCATGTGTATTACGCGCTGGGCGCAAGCTGGGCTATCCAGCGCAGCAAGCGCTTA 300
QY 259 CACCGCGTGGCCGGAAGGCTG-----GAGAGCTTCCCCCGAGCTTCCGC 303
DB 301 CACGATGCGCACCGTATCGCCCGCACCAATATGATGTGTGCCGACATATTCAG 360
QY 304 GAGCTTGGGGGCTTCTGTGCTCGGAGCTTACACCGCGGCGGCGGCTCCATCGCC 363
DB 361 ATCTGTGTACCTTGTCCGGGCGTCCGAGCTACACCGCGCGCGGTGGCTTTCCT 420
QY 364 TTCGGGAGCGGGTGGCGGCGGTGACGGAACGTCCGAGGCTCTCTCCGCTTT- 422
DB 421 TACCGCCAGCGGGTGGCGGTGTGACACCAATGTGCGCGCGTGTGCGCCGCGCTT 480
QY 423 -----CGCCCGGAAAGCCCAAGAGAGCTTTTCCCTCGCCCAAGGC 471
DB 481 CACGCGCGCGCGACCGCGGTGCGCATCGGTGCGCGCGACCAACCGCGAGCTCTTGGCG 540
QY 472 CTCCTCCCGAGGCGGTGGAACCGGGGGGTGTGAAACAGCGCCCTCATGAGCTCGGGCC 531
DB 541 CTGTGCGCGACCGCGAGACGGCGCTGAATTTTCGTGCGGCTGTATGAGTTGGGTGG 600
QY 532 ACGGTCTGCTGCGGAACCGCGCGGTGCGGGGCTGCGCTTAGGGGCTTCTGCGCG 591
DB 601 ACGGTGTGACCGCGCGGACACCGCGGTGTATGCGCGTGTGGAATGGTGGCATGG 660
QY 592 GGGGAAGAGCGCCCGGCGCTACCCCGCGCCAGGAAGCGCGGCGAAGAGAGCGC 651
DB 661 CGGATGTGCGGTATATCCGCGCTCGGACGCTCC---GCCGCGCGGGGCGAGGCTTACAC 717
QY 652 CTCGTGCGCTGCTCTCTCTCGGGCGG 678
DB 718 GGAACCGACCGCAAGTCCGCGAGCG 744

RESULT 15
AB279955
ID AB279955 standard; DNA; 1312 BP.
XX
XX AC AB279955;
XX
XX 19-MAY-2003 (first entry)
XX
XX Mycobacterium tuberculosis muty nucleotide sequence SEQ ID NO:21.
XX
XX Mycobacterium tuberculosis; mutR2; alkA; ogt; RV3908; mutY; RV3909;
KM detection; multidrug resistance; multiple drug resistance; MDR;
XX infection; gene; ds.
XX
XX Mycobacterium tuberculosis.
XX
XX MO2003015662-A2.
XX
XX 27-FEB-2003.
XX
XX 14-AUG-2002; 2002MO-EP009679.
XX
XX 14-AUG-2001; 2001US-0311824P.
XX
XX 21-AUG-2001; 2001US-0313523P.
XX
XX (INSP ) INST PASTEUR.
XX
```

```
PI Giequel B;
XX
XX WPI: 2003-256711/25.
DR
XX
XX Predicting the epidemic character of a Mycobacterium tuberculosis isolate
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate
PT by detecting an alteration in the DNA repair system of the isolate.
XX
XX Disclosure; Fig 6A; 83pp; English.
XX
XX The present invention describes a method for predicting the epidemic
CC character of a Mycobacterium tuberculosis isolate and/or a selective
CC advantage to be maintained in the host and/or the acquisition of multiple
CC drug resistance (MDR) by the isolate comprising detecting an alteration
CC in the DNA repair system of the isolate. Also described: (1) detecting a
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
CC polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)
CC detecting in a patient infected by Mycobacterium tuberculosis a higher
CC risk of being unable to eliminate the bacillus or of developing MDR
CC tuberculosis. The method is useful for predicting the epidemic character
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to
CC be maintained in the host and/or the acquisition of MDR by the isolate.
CC The present sequence represents a M. tuberculosis muty nucleotide
CC sequence, which is used in the exemplification of the present invention
XX
XX Sequence 1312 BP; 173 A; 471 C; 457 G; 211 T; 0 U; 0 Other;
SQ
Query Match 8.9%; Score 86.6; DB 7; Length 1312;
Best Local Similarity 50.2%; Pred. No. 2.7e-06;
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;
QY 22 CTCCTGCTGTGATCGGGAAGACCGCGCCCTCTCCCTGGCGGGG--GAGAAGAC 78
DB 261 CTTCTCGCTGTGTATCAAGCATCGACCGGATCTGCTTGGCGAGACCGCGTGTACG 320
QY 79 CCTTACCGCGTCTGTCTCGAGTCTTCTGTGAGAGACCGCGGTGTGAGACGCTTC 138
DB 321 CCGTGGAGATCTGTGTACGAGATTATGTGAGAGACGCGCGCGCGGTGTG 380
QY 139 CCTATTACCGCGCTTGTGAGCGCTTTCACACCTGAAAGCGCTTGGCGGCTTTC 198
DB 381 GCGATCTGGCGGACGTGGTGGCGGTGGCGGCGCGCGCGCGCGCGCGCGCGCG 440
QY 199 CTGGAAGAGTCTTAGGCTGTGCAAGGGCGGCTTACTACCGCGCGGGAACACTC 258
DB 441 ACCGCGATGTATTACCGCGCTGTGGCGCAAGTGGCTATCCAGAGCAACGCTTA 500
QY 259 CACCGCTGGCCGGAACGCTG-----GAGAGCTTCCCCCGAGCTTGGCC 303
DB 501 CACGATGCGCGCACCGTATCGCCCGGACCAATATGATGTGTGCCGAGATATGAG 560
QY 304 GAGCTTGGGGGCTTCTGTGCTGCGGCTTACACCGCGGCGGCGGTGAGCTTCATGCC 363
DB 561 ATCTGTGTACCTTGTCCGGGCGTGTGAGACTACCGCGCGCGGTGTGCTTTCCT 620
QY 364 TTCGGGAGGCGGTGTGCGCGGTGAGACGGAACGTCCGAGAGGCTCTCTCCGCTTT- 422
DB 621 TACCGCAAGCGGGGTGGCGGTGTGACCAATGTGGGCGGTGTGGCGCGCGCTT 680
QY 423 -----CGCCCGGAAAGCCCAAGAGAGACTTTTTCCTTGGCCAGGCG 471
DB 681 CACGCGCGCGCGACCGCGGTGCGCAATCGGTGCGCGGACCAACGCGCATCTTGGCG 740
QY 472 CTCCTCCCGAGGCGGTGGAACCGGGGATGTGAAACAGGCGCTCATGAGCTCGGGCC 531
DB 741 CTGTGCGCGACCGCGAGAGCGGCGCTGAATTTTCGTGCGCTGTATGAGTTGGGTGG 800
QY 532 ACGGTCTGCTGCGGAACGCGCGGTGTGCGGGCGCTCCCTTAGGGGCTTCTGTGCGG 591
DB 801 ACGGTGTGACCGCGCGCACACCGGTGTGCGGTATATGCCGCTGTGACTGTGTGCA 860
QY 592 GGGGAAGAGCGCCCGGCGCTACCCGCGCCCAAGAGAGCGCGCGGAGAGAGCGC 651
```

Db	861	CGGCA	TGCCCG	GTTAT	CCGCC	GTCG	AGCG	GTCC	---	GCCG	CGCC	GGGG	GCAG	GCTT	ACAC	917
Qy	652	CTCG	TGCC	CTCG	CTCT	CTCG	GGCG	678								
Db	918	GGAC	CGAC	CGCC	CAAG	TCCG	CGGAC	CG	944							

Search completed: March 4, 2004, 04:19:18  
Job time : 366 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 4, 2004, 03:56:41 ; Search time 2811 Seconds

(without alignments)  
15033.603 Million cell updates/sec

Title: US-09-938-901-1

Perfect score: 975  
Sequence: 1 gtgtgagggccgcgcgaagc.....tagtccccctccgcgaagca 975Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

```
GenEmbl.*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_un:*
13: gb_vl:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_mu:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rdd:*
36: em_hcg_mam:*
37: em_hcg_vtc:*
38: em_sy:*
39: em_hugo_hum:*
40: em_hugo_mus:*
41: em_hugo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	975	100.0	975	6	BD174408
2	176.4	18.1	1089	1	BD174408 DNA resto
3	176.4	18.1	11378	1	AF377342 Deinococc
4	163.2	16.7	8249	1	AE543480 Deinococc
5	157.8	16.2	303642	1	AE543480 Uncultured
6	144.2	14.8	190050	1	AE016923 Chromobac
7	135	13.8	257701	1	AE644022 Ralstonia
8	135	13.8	346259	1	AE644035 Bordetell
9	135	13.8	346287	1	AE644050 Bordetell
10	132.2	13.6	293050	1	SC0939116 Streptomy
11	130.2	13.4	346547	1	AP003012 Mesorhizo
12	129.4	13.3	1106	1	AF121797 Streptomy
13	127.8	13.1	11283	1	AE005710 Caulobac
14	124.4	12.7	11034	1	AE005710 Rhodospir
15	123.8	12.7	11786	1	AE005065 Halobacte
16	123.8	12.7	11786	1	AE004927 Pseudomon
17	118	12.1	300531	1	AE016932 Bacteroid
18	116.6	12.0	10923	1	AE014358 Brucella
19	116.4	11.9	297700	1	AE005944 Brucella
20	115	11.8	9829	1	AE009581 Brucella
21	114.8	11.8	349970	1	BX571659 Sequence
22	113	11.6	1551	6	AR388059 Sequence
23	112.4	11.5	299800	1	AP005040 Streptomy
24	111.6	11.4	303438	1	AE017212 Geobacter
25	110.4	11.3	310613	1	AE016875 Pseudomon
26	109	11.2	299950	1	AP005372 Thermosyn
27	108.8	11.2	286550	1	SMES91785 Simorhizo
28	107.8	11.1	300511	1	AE016775 Pseudomon
29	107.2	11.0	302988	1	AE017176 Porphyrom
30	105.8	10.9	299700	1	AP006571 Geobacter
31	105.4	10.8	13816	1	AE019335 Yersinia
32	105.4	10.8	208050	1	AJ414145 Yersinia
33	104	10.7	1688	9	AB032925 Homo sapi
34	104	10.7	1692	9	AB032929 Homo sapi
35	104	10.7	1721	9	AB032924 Homo sapi
36	104	10.7	1756	9	AB032928 Homo sapi
37	104	10.7	1757	9	AB032923 Homo sapi
38	104	10.7	1759	9	AB032927 Homo sapi
39	104	10.7	1821	9	AB032922 Homo sapi
40	104	10.7	1824	9	AB032921 Homo sapi
41	104	10.7	1838	9	AB032926 Homo sapi
42	104	10.7	1844	9	BC003178 Homo sapi
43	104	10.7	1854	9	AB032920 Homo sapi
44	104	10.7	1869	9	HSU63329 Human mult
45	102.8	10.5	348764	1	BX569689 Synechoco

## ALIGNMENTS

RESULT 1	LOCUS	BD174408	975 bp	DNA	linear	PAT 18-MAR-2003
DEFINITION	DNA restoration enzyme gene.					
ACCESSION	BD174408					
VERSION	BD174408.1 GI:29120096					
KEYWORDS	JP 2002247985-A/1.					
SOURCE	Thermus thermophilus					
ORGANISM	Thermus thermophilus					
REFERENCE	1 (bases 1 to 975)					
AUTHORS	Kuramitsu, N. and Yokoyama, S.					
TITLE	DNA restoration enzyme gene					
JOURNAL	Patent: JP 2002247985-A 1 03-SEP-2002;					



QY 126 GAGACAGCCCTCCCTATTACCGCCGCTTCTGAGCGCTTCCACCTGAAGGCCCT 185  
Db 186 GGGCGGGGGGCTGGGGTACTACGAGCGCTTCTTGAGCGCTTTCGCGGGCGGCCCT 245  
QY 186 GGGCGGGGCTTCCCTGAGAGAGGCTCTTGAAGGCTTGAGGAGGGCGGCTTCAACCGCG 245  
Db 246 GCGCGCGCGCGCGCGAGAGCGCGCTGCTCAAGGCTTGAGAGGTTGTGCTACTACGCGCG 305  
QY 246 GGGGGAACACCTTCCACCGCGCTGCGCCGAGAGGCTGAGAGAGC-----TTCCCGCGAGCTT 299  
Db 306 TGGCCCGACCTGACCGGGGAGCGGCGCATCATGACGAGCAGGGGTTTCCGCAAGACTA 365  
QY 300 GCGCGAGCTTGGGGGGCTTCTGCTCTGCGGCTTACACCGCGCGCGGCTGCTCAT 359  
Db 366 GCGCGGGTGGCTGGCTGGCTGGCGCGCTGCGCGCTTACACCGCGCGCGGCTGAGCACT 425  
QY 360 GCGCTTGGGGGAGCGGGTGGCGGGTGAACGGGAGCGTCCGAGGCTCTCCGCGCT 419  
Db 426 GCGGCTGGGGGAGCGCGGAGCGGTGAACGAGCGGAGCGGCTGCTGCTGCGCGCT 485  
QY 420 CTTCGCGCGGAGAAAGCCCAAGAGAGAGAGCTTTTCCGCTCGCCGAGGCTCTCTCC 479  
Db 486 GCGGGGAGAGGCTACCGCTCCGCAAAATGGGTGACAGGAACAGGC---TGACGCTTGTCT 542  
QY 480 GAGAGGCTGAGACCGGGGGGTGTGAACAGGCGCTTATGAGCTCGGGGCTACGCTGTG 539  
Db 543 GAGACCGGCTGAGACCGGGGTGCTGTGAACAGGCGCTTATGAGCTCGGGGCTACGCTT 602  
QY 540 CCGGCGAAACGCGCGGCTGCGGGGCTGCGCGGCTTCTGCGCGGGGAA--- 596  
Db 603 GGTGCGCAAGTCCGCGGCGGTGACCGCTGCGCGCTACGCGCACTGTGCGGCTTACCA 662  
QY 597 ---GAGAGCCCGCGGCGCTACCGCGCGCCAGAGAGCG---CCGGCGAGAGAGAGAGCG 650  
Db 663 ACTGGGTCAGCGCGGTACTTCTGCGCCCAAGCGCGCGCGCGCGAGGTCG 722  
QY 651 CTTGCTGCTGCTCTCTCTCTGCGGCGAGAGGGGCTGACCTGGAAGGCTTGAAGGCG 710  
Db 723 GCGCGTGGCGCTCTGTGATCGGGAGCGCGGAGTACCGCTGTGAAAAAGCGGAGGCTC 782  
QY 711 CTTCCAGGCG 720  
Db 783 GCTGCTGGCG 792

RESULT 3  
AE002060/c 11378 bp DNA linear BCT 22-NOV-1999  
LOCUS Deinococcus radiodurans R1 section 197 of 229 of the complete  
DEFINITION chromosome 1.  
ACCESSION AE002060 AE000513  
VERSION AE002060.1 GI:6460082  
KEYWORDS  
SOURCE Deinococcus radiodurans R1  
ORGANISM Deinococcus radiodurans R1  
Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
Deinococcaceae; Deinococcus.  
REFERENCE 1 (bases 1 to 11378)  
AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,  
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,  
Mofat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,  
Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.  
TITLE Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans R1  
JOURNAL Science 286 (5444), 1571-1577 (1999)  
MEDLINE 20036896  
PUBMED 10567266  
REFERENCE 2 (bases 1 to 11378)  
AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,  
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,  
Mofat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.

FEATURES  
source  
TITLE  
JOURNAL  
Direct Submission  
Submitted (08-NOV-1999) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1..11378  
/organism="Deinococcus radiodurans R1"  
/mol\_type="genomic DNA"  
/strain="R1"  
/db\_xref="taxon:243230"  
/chromosome="1"  
gene  
118..879  
/gene="DR2274"  
118..879  
/gene="DR2274"  
CDS  
118..879  
/gene="DR2274"  
/note="similar to PID:1001216 PID:1001276 percent  
identity: 51.02; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAPI1822.1"  
/db\_xref="GI:6460083"  
/translation="MTEQAQYRDPRLVLYDRINRMGADDFPLANETPGCRLLDL  
GCGTGRITTLARAGQVGTGEPARSLAEVAGRPGEACVWLOGSQADPTAFADLC  
LMTAHVAQVEVEDDAMQDVLQIHRLALVPGGRSLFDMRDPAAAMDYDSDAGKERTLT  
LLDGSAAETGCDVLDAVGATVSGPVVHFAETRELPQDVLTSRRLFRREDELRLAS  
LQAGFAVBOVYGMGEPIVGTGCGELVVAARRPG"  
840..3032  
/gene="DR2275"  
840..3032  
/gene="DR2275"  
CDS  
840..3032  
/gene="DR2275"  
/note="similar to SP:056243 PID:1325924 percent identity:  
79.97; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="exonuclease ABC, subunit B"  
/protein\_id="AAPI1823.1"  
/db\_xref="GI:6460084"  
/translation="MRTTGRGAGQAGVTRLHPFRLPSSLRGCRKLLADVOTLEPRL  
CVCLAHRCRTTLEVMALKRSEKPSGSDOTFAISLYDGLSEGLRYTLIGATGTCT  
YSMAKYLEVQRPALINAPNKILTAQLASFREFFPAAVEFLISTDYIOPETAVFG  
KDLFIERDASVNOEIERLRHSTRLTRDITIVASVSCITYGADPKETALNAIVK  
KGVMPREDLGRILVMQYERNDIELMPGFSGVGEVSVTIGADPKETALNAIVK  
ERISVHPLTGERLGLDADATVVYPAKHVVSAGNIBRAGSIOOELBERLEYFHSVGK  
LLEAQLKERITLDLEMLKVLGYCSGIESNSRHIDGAPGATPYTMIDYRDPDITRI  
DESHVTPQIGMANGRARKQTLVDYGFRLPSAMDRPLNPFSEFSLKGTQYFVSAT  
PGPFENQSDSVADQIIRPTGLADPEVTRIPGQILDLGVRERTAAGEBTLVTL  
TKRMSEDLTLEYLLEKGVKARVHSDIDSVERQVIRDLRGLGHDVLVGLNLRGLDL  
PEVSLVALDADKGLFASDRALIGTIGRAARVNGEIVLADVYTPAMFAMDENLR  
RREKOLAIVNEHGIPTPTVTKGVADVIRGELEGEISEVNSGDRSLTQLTDLIED  
MMASEMDPEKASLRDQIRALIAKLGKEFPGATYPOGKARRKGR"  
3035..3184  
/gene="DR2276"  
3035..3184  
/gene="DR2276"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAPI1832.1"  
/db\_xref="GI:6460093"  
/translation="MGRGAKMLPAJLALIFLVVALVWMTGPTAVVLAALCPAAVLS  
ASGRC"  
complement (3287..4006)  
/gene="DR2277"  
complement (3287..4006)  
/gene="DR2277"  
/note="similar to GB:AE0000782 percent identity: 66.67;  
identified by sequence similarity; putative"

gene  
CDS  
/codon\_start=1  
/transl\_table=11  
/product="amino acid ABC transporter, permease protein"  
/protein\_id="AAFI1824.1"  
/db\_xref="GI:6460085"  
/translation="MSRSKPGGMDLTGTQTAISGDYPOQLLSGLGTLAVSICALV  
USVVTGALGVRYVRVPLBPGIVGVEVGRIPILVLSIYFGLPTLGLTAGEP  
AAVLAIGLSAATSEIVRGGLGSGVPAQIEASRLSGOALRFVPLPQAMVAP  
ALGNEFVLSLSSLASAVTQDELFSQGRITNATYRQFEYVAVALVIFLFTLIR  
AVRALEHRLSRQSLPERRVI"  
/complement(3993..4811)  
/gene="DR2278"  
/note="similar to SP:P10344 PID:41569 GB:U00096  
PID:1651365 PID:1651371 percent identity: 53.85;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="amino acid ABC transporter, periplasmic amino  
acid-binding protein"  
/protein\_id="AAFI1825.1"  
/db\_xref="GI:6460086"  
/translation="MGLVRNVGRILNRVSGMKRUSALLVLAISCSGASAYAV  
KKGTLVLGDPFAPFEFKSGDGVQGFIDIAVAHADLVLRIRVPGALMPO  
AATSGRVVAMSGITITPERATVTFPSQYSAQVIVRSNGPQFTMPADVKKT  
GUAQNTGOVAEQTLKPRGATIKYVDPAAGLADVSRIALIGDAFTVDLKKRL  
EGQFDKAGNVAVMEDYGMVFRKGSDLAAANRTLAALKRSGEYOKLNMIVQK"  
4846..6141  
/gene="DR2279"  
4846..6141  
/gene="DR2279"  
/note="similar to GB:AL009126 percent identity: 59.28;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="alcohol dehydrogenase, zinc-containing"  
/protein\_id="AAFI1826.1"  
/db\_xref="GI:6460087"  
/translation="MGERKCGWPKPVFLHLSLSPGAGSTKAVVHMGIDRLDVP  
EPKTEAPDAIVRLTASALIGTDLHFRGMSDMVPRDTLIGHGAVGIVEVGEVNE  
VPGDRVIVPSASGCGYPCPRBEGNTSQCDNANPAGPAGFSGSLGMDQAE  
KVRVYVANSLSVLKLDNVSDQAIMSDIPTAYFGAEIGVTVGSVVVVLGGGACV  
FSIISAKLIGTRVIAVDRDLDMKMKGAIVINEKEDPVEAKYKTLGGLADVC  
IDAVGVADHNAHAPAKPDAKEQKEQVEQVAVADKPERDQMTPGDAPSAASMA  
IEMTKAGQIGIIGVSPQMTYPIGKAMKNLITRNQCNHRAIIPGLVDVAAGV  
DPAVLTEHHLGSAIEFAFDKRPQGMIKVLELPQAG"  
complement(6214..7704)  
/gene="DR2280"  
complement(6214..7704)  
/gene="DR2280"  
/note="similar to GB:X76083 SP:P41006 PID:43121 percent  
identity: 70.23; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="uracil permease"  
/protein\_id="AAFI1827.1"  
/db\_xref="GI:6460088"  
/translation="MHLAVMARAGQSSSSAWGLDARDFYREARALRARRPMSATL  
RMTSSSEKATAEACFCSPSPAHYTPMTQTAPELDLPQRRVVLGLQHSIM  
RGSTIVLVPLVINGSPVALFAGVATLIFHLTRGQVIFLSSFAFLAPALVIKEM  
GPAANAAGLIAAGLMYVPSGIVKIFGTERLLRPVPVTPVIVIGLGSVAVQ  
AKTNGLALVTLAALIASITGRGLFRMIPILIGVTVGVVLAALTGQVLAADIAA  
APLGLPPEHGPALDMRAVAIIAPVAIVTFEHVGDVNVGRVGNKFLNPGLSRTL  
LADGIANSSAVMGPAATYANTGVALTTRVDPVIOGAVFALLFGSPKLAAY  
LOGLPQGVIGVSLIFGMIASVIGIRLAEQVDFASHRNLIIVSLIIVLGLGASFP  
LVNAGTKLEIHMALALVGIIVANLLRPRPEISKADGMPGVG"  
7746..8240  
/gene="DR2281"  
7746..8240  
/gene="DR2281"  
/note="identified by Glimmer2; putative"

gene  
CDS  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI1833.1"  
/db\_xref="GI:6460094"  
/translation="MGACQAKAEVLTPPPPOSTRTLYPMNAKGDLLRLVSLGICTP  
TEEAHGPHERTHPVKWSGVIAIAEGRTTKDARLAEALRLDSSDAPVP  
AAPATPIAQSSPWPVIYQVLAEAAMEFARBDATLDEVARDAGRFRREBLADLGH  
GPA"  
8243..8614  
/gene="DR2282"  
8243..8614  
/gene="DR2282"  
/note="similar to SP:P45465 PID:606088 GB:U00096  
PID:1789538 percent identity: 56.57; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI1828.1"  
/db\_xref="GI:6460089"  
/translation="MPTSRPRGHOGAABARAAALHESLGREIVARNYRIPGEIDL  
VSRPGLTVFTEVRORARVGSALDSVTPKRLMHRAALEVLTREGRDDLPCRL  
EVLIRGEADVTGELRLAVEG"  
complement(8625..9431)  
/gene="DR2283"  
complement(8625..9431)  
/gene="DR2283"  
/note="similar to SP:Q55282 percent identity: 66.27;  
identified by sequence similarity; putative"

Query Match 18.1%; Score 176.4; DB 1; Length 11378;  
Best Local Similarity 58.4%; Pred. No. 9.1e-15;  
Matches 391; Conservative 0; Mismatches 261; Indels 18; Gaps 4;

gene  
CDS  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI1833.1"  
/db\_xref="GI:6460094"  
/translation="MGACQAKAEVLTPPPPOSTRTLYPMNAKGDLLRLVSLGICTP  
TEEAHGPHERTHPVKWSGVIAIAEGRTTKDARLAEALRLDSSDAPVP  
AAPATPIAQSSPWPVIYQVLAEAAMEFARBDATLDEVARDAGRFRREBLADLGH  
GPA"  
8243..8614  
/gene="DR2282"  
8243..8614  
/gene="DR2282"  
/note="similar to SP:P45465 PID:606088 GB:U00096  
PID:1789538 percent identity: 56.57; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI1828.1"  
/db\_xref="GI:6460089"  
/translation="MPTSRPRGHOGAABARAAALHESLGREIVARNYRIPGEIDL  
VSRPGLTVFTEVRORARVGSALDSVTPKRLMHRAALEVLTREGRDDLPCRL  
EVLIRGEADVTGELRLAVEG"  
complement(8625..9431)  
/gene="DR2283"  
complement(8625..9431)  
/gene="DR2283"  
/note="similar to SP:Q55282 percent identity: 66.27;  
identified by sequence similarity; putative"

66 GGGGAGAAAGAACCTTACCGGCGCTGCTCGAGAGTCTTCTGAGAGAACCCGGGT 125  
11165 GGGCGGGCGGCACTTACCGGCGCTGCTCGAGAGTCTTCTGAGAGAACCCGGGT 11106  
126 GAGAGAGGCGCTTCCCTTATTCAGCGCGCTTCTGAGAGTCTTCTGAGAGAACCCGGGT 165  
11105 GGGCGGGGCGCTTGGGATCTAGAGCGCTTCTGAGAGTCTTCTGAGAGAACCCGGGT 11046  
186 GGGCGGCGCTTCTGAGAGTCTTCTGAGAGTCTTCTGAGAGAACCCGGGT 245  
11045 GGGCGGCGGCGCGGAGAGCGGCTGCTGAGAGTCTTCTGAGAGAACCCGGGT 10986  
246 GGGGGAACACCTCCAGCGCTGCGCGGAGAGTCTTCTGAGAGAACCCGGGT 299  
10985 TGCCTCCGCACTGCAACCGGAGAGCGGATCATGACGAGAGAGGTTTCCGCAAGACTA 10926  
300 CGCGAGCTTGGGGGCTTCTGCTGCTGAGCGCTTACCGCGCGCGGCGGCTCCAT 359  
10925 CGCGGAGTGGCTGAGGCTTCCGCGGCGGCGGCGGCGGCGGCGGAGAGCT 10866  
360 CGCTTCCGGGAGCGGAGTGGCGGCGGAGTGGAGAGAGTCTTCCCGCT 419  
10865 CGCGTGGCGAGCGCGGAGCGGAGTGGAGAGAGTCTTCCCGCT 10806  
420 CTTCGCGCGGAGAGCGGAGTGGAGAGAGTCTTCCCGCT 10749  
10805 GCGGAGAGAGTCCGCTCGACAAATGGGTGAGAGAGAGTCTTCCCGCT 10749  
480 CGAGGCGTGGAGAGCGGAGTGGAGAGAGTCTTCCCGCT 539  
10748 GGAACCGCGTCCGAGCGGAGTGGAGAGAGTCTTCCCGCT 10689  
540 CCGCGGAGAGCGGAGTGGAGAGAGTCTTCCCGCT 596  
10688 CGTCCCAAGTCCGCGGAGTGGAGAGAGTCTTCCCGCT 10629  
597 ---GAGGCGCGGAGTGGAGAGAGTCTTCCCGCT 650  
10628 ACTGGGTGAGCGGAGTGGAGTCTTCCCGCT 10569

Oy 651 CCTGTCGCTCTCTCTCTCGGGCGGAGGGGTGCACTTGAAAGCTTGAGGGCG 710  
Db 10568 GGGCTGGCTCTCTCTGATCGGGACGCCAGTACGCCCTGCTGGAAGACGGAGGGCTC 10509  
Oy 711 CTTCAGGGC 720  
Db 10508 GCTGCTGGC 10499

RESULT 4  
AF543480 8249 bp DNA linear BCT 07-MAR-2003  
LOCUS DEFINITION  
AF543480 uncultured bacterium plasmid pAK106 Soda (soda) gene, partial cds;  
Grea (grea), Muty (muty), putative aminopeptidase, and putative  
AF543480 response regulator genes, complete cds, and unknown genes.  
AF543480.1 GI:28875482

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
SOURCE

1. 8249  
/organism="uncultured bacterium"  
/mol\_type="genomic DNA"  
/isolation\_source="soil"  
/db\_xref="taxon:77133"  
/environmental\_sample  
/plasmid="pAK106"  
/note="anonymous microbial sequence conferring formation  
of carbonyls from short-chain polyols on Escherichia coli"  
1. 618  
/gene="soda"  
/note="putative superoxide dismutase"  
/codon\_start=1  
/transl\_table=1  
/product="Soda"  
/protein\_id="AA059964.1"  
/db\_xref="GI:28875483"  
/translation="DLAYPEGEFRFPPEAGLSATLATHIELYRGVVAQRELLGL  
ADBSLRKAGAVTRPREAVSRKMSFEINGVTLHWEYEQLAGRSGLRVPDSSVAAGS  
MDVAFGGFDGRADVALELATTRGIVVAAMDASQGLVINMADLHLLLPAGHIVF  
VLDMEHAYMDDFGPKRGAVSKFLENTDVEVEARFARANGA"  
complement (673..2466)  
/gene="grea"  
/note="grea"  
/gene="grea"  
/note="putative transcription factor"  
/codon\_start=1  
/transl\_table=1  
/product="Grea"  
/protein\_id="AA059965.1"  
/db\_xref="GI:28875484"  
/translation="MRHADGLOGASALRPAGADAGAOVHQAALVRDALLRGGQGLG  
LRPQRTFVCGPREVATVTECGHPFDVALIEDGPHLAERGDGRCGAPTCGRGQH  
LAAREBELAAMHHRLRTAVOVAARAAVAPREHVVORRGORLHVGTTOEPGV  
VVDHGLRLIOLHILGQPHAVRVICALPGQVAMPLPAPHNAGALOLHSSGSVSR  
RGACRMFGPCTRDVFGGGHQAIVHLLQGFAPLARKFLGADLLQVTRGDLADLQCLAR

gene  
CDS  
1834..2877  
/gene="muty"  
/note="muty"  
/note="putative adenine glycoylase"  
/codon\_start=1  
/transl\_table=1  
/product="muty"  
/protein\_id="AA059966.1"  
/db\_xref="GI:28875485"  
/translation="MKLESLASRYVRWQROHGRHGLPMQGTODPYRVMLEVMLOQTQ  
VVTVIDYARFLQRPDVALAALPDDVLAALMGLYGRARHLHCAQAVVQHG  
EPPSSSEMLTLPEIGRSTAAATAAFCFGRVAIIDGNVRLTRALGFGPDLSPAN  
ERALMAOQTLPLAQGTITAYTQGLMDGAGICTLRBPOCAACPLQPCVAAHOGRPA  
YPVTRRLKRGQRAVVMVLMHGLQVAVRPEQGVAGLSLDEMPADSLGTLAA  
WPGHEWLPVPALTHFDWTLHPLRWQLPAPVAPARRKAVEQALPTGRWFTQSEALA  
ASLPAPYRKLAC"  
complement (2904..3335)  
/note="Orf6; hypothetical protein"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1  
/product="unknown"  
/protein\_id="AA059967.1"  
/db\_xref="GI:28875486"  
/translation="MKRTTFLTAIALSATSAGSVAPQTTDTGSLRDKAKCTYLRD  
KDVAESRCFQDNCALVPPPMADGAGKLTIVARKGGQQAAMGMPLYTPAGDA  
KVGDTSGSGGWHVTKAAGDKPKBRATGNADPYKY"  
3742..4827  
/note="Orf7; hypothetical protein"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1  
/product="unknown"  
/protein\_id="AA059968.1"  
/db\_xref="GI:28875487"  
/translation="MRHGCSAAHCGSSSSAMVPAHRAVACQARLCSAGSSSP  
TSSRDAAATTPPYVSQAQAPVPGDQOABSPATQVAVVAASGMAAAGSGRRAD  
HSPASASHPRSVCVAAAMATQARPRRTSRPGAPSSARPVQDCTAPIDV  
FSHAGPALRLRAPRGVWRRIIDQAPPAQPPGPGFSFKHPAACNVVIGOPAYERK  
DRLPVLRLRPPQTAQHRQAAGSLSPSPSCVWRDAGIVSAQPSRSGSIVRSARR  
ARRSMARLTCCOPGHRPCDREKQABERGFLVVGPARGAPAGEWVRCRSRMSRRRA  
WPAWSRSRPRVPRPCSRRLRRPPWR"  
complement (5521..5868)  
/note="Orf8"  
/codon\_start=1  
/transl\_table=1  
/product="putative aminopeptidase"  
/protein\_id="AA059969.1"  
/db\_xref="GI:28875488"  
/translation="MNALVETNDSGLNDIRGQVTPAQVVDALRAARSGFVAGSVG  
AGCTGIAFGKKGIGITSSRVAVTAGSSSYTVLVQSNYGDVLVWAGVPVAGSSPPGGA  
RTRVPVRPARA"  
5874..6179  
/note="Orf9"  
/codon\_start=1  
/transl\_table=1  
/product="putative response regulator"  
/protein\_id="AA059970.1"  
/db\_xref="GI:28875489"  
/translation="MGQPLLDTVARRAQNGBRDLAARENAVITVLARVEKIVSK  
EALIQAVAMGSELTPALISVYSRLPAKLEPAGVKIRTVGFGYMLBEFRKKAQR"  
complement (6591..7028)  
/note="Orf10; hypothetical protein"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1

```

/product="unknown"
/protein_id="AAO53971.1"
/db_xref="GI:28875490"
/label="1"="MPRPPTSGAREGCDGKRGHTEFAVEALIGRALPDEVETMTGQD
ALVAKHKEVCEYASHKSAEENSLVYMAFAPADALVDPDPATYTYLRKHLYLNHPCVLVR
LIRLRDALRDLVAVCAHLRVNVOAKRSKSDIGSGSRPLDRR"
complement(778..>8249)

```

```

/product="unknown"
/protein_id="AA059972.1"
/db_xref="GI:28875491"
/translation="IGRRPPTAGCAATPEDAWVSRSGRTDRPPTGCAITGDM
RQHRRPLAATPGSPSPHPQAPDAGSPPEAAVHAGGADDDDEFDRRPLAVFV
VARGAASGGRRRRRCDDHDAFEDAEHNHRSHSAELQGRGRGQVSSGE"

```

**ORIGIN**

Query Match	16.7%;	Score 163.2;	DB 1;	Length 8249;
Best Local Similarity	54.7%;	Pred. No. 5.8e-13;		
Matches 408; Conservative	0;	Mismatches 308;	Indels 30;	Gaps 3

0y 2 TGGAGGCGCTGGCGGAAAGCCCTCTCTCGCTCGCTACCGGGAAAACGCGCGCGCC---CTCC 58  
 |||||  
 Db 1841 TGGAGAGCCTGGCGCAGCCCGCTGTCTCGCTCGCAGCGGCAACATGCGCCGCCACGGCGTGC 1900

0y 59 CCTGCGGGGAGAAAGACCTTACCGCGTCTGGTCTCCGAGGTCCTTGTGCACAGA 118  
 1901 CCTGCGGGGACACAGAGCCCTGATCCGCGTGTGCGTCTCCGAGGTGATGCTGCACACAGA 1966

QY 119 CCGGGGTGAGACAGACCCCTCCCTCCTATTACCGCCGGCTTTCGAGAGCCCTTCCACCCTGA 178  
 Db 1961 CGCAGGTGCTCACCGCTGATGCACTACTACGCCGGCTTCCTGCACGTTTTTCCGAGCTGC 202

Accession	Sequence	Length
Qy	AGGCGCTGCGCGCGCGCTTCCCTGGAGAGGTCCTTAGGGTCGGCAGGGGGGGGGGCTACT	238
179	AGGCGCTGCGCGCGCGCGCTTCCCTGGAGAGGTCCTTAGGGTCGGCAGGGGGGGGGGCTACT	238
2021	AGGCGCTGCGCGCGCGCGCGCTTCCCTGGAGAGGTCCTTAGGGTCGGCAGGGGGGGGCTACT	208
Db		

[illegible]

Oy	AGCTTCCCGCAGCTTCGCCCAGCTTCGGGGCTCTGTGTTCTGGACCTTACATCCCG	343
Db	AGTTCGCCGCAGCACTGAGTAGTCTGTGCTGCCTGCCTATCCGCGCTCTCACACCGCG	220

Qy 344 CGGCGGTGGCTTCATTCGCTTCGGGAGCGGCTGGCGCGGATCCGGA 403  
 Db 2201 CAGCCATCGCGCGCTTGTGCTTCGGGCGAGCGGGTGGCCATCCTCGATGGCAACGTCAAC 226

[illegible]

Oy 464 CCCAGGACCTCTCTCCAGGGGCGAGACCCGGGGGAGAGAAC TTGG 232  
 Db 2321 CGCTGTGGGCGCAGGCCAGACTTGTGGCCCGCGCAGGGGCATCACGGCCTTACACGCGAG 238

OY      512    CCCCTAATGAGACATCGGGCCCAAGATCTGCCTCCTAAATCATCTCTTCTTC

Db      2381   GCGTGATGACCTGGGGCGCGGCATCTGCACCCTTGCGCCGGCGCAGTGCCTGTGCC

Db	Y	QY	2441	250
CCTTGC	UCC	UCC	UCC	UCC
CAACCC	UCC	UCC	UCC	UCC
CGTCTG	UCC	UCC	UCC	UCC
CGCGCAT	UCC	UCC	UCC	UCC
CGCAGGCG	UCC	UCC	UCC	UCC
CGCCGCG	UCC	UCC	UCC	UCC
CCGACG	UCC	UCC	UCC	UCC
CTTACCC	UCC	UCC	UCC	UCC
CGGTAGA	UCC	UCC	UCC	UCC

[illegible]

QY 692 TGGAAAGGCCTGAGGGGCTCTCCAG 717

Db 2561 AGGTCTGGCTCGTGGCCCGAG 2586

RESULT 5	AE016923/c	LOCUS	AE016923	303642 bp	DNA	linear	BCT 14-OCT-2003
DEFINITION	Chromobacterium	violaceum	ATCC 12472	section 14 of 16	of the		

ACCESSION	AE016923	AE016825
VERSION	AE016923.1	GI:34104942

**SOURCE ORGANISM**

## REFERENCE

**TITLE**

JOURNAL  
MEDLINE

## REFERENCES

TITLE  
JOURNAL

## FEATURES

gene

CDS

303642 bp DNA linear BCT 14-OCT-2003

Genome: AEO16825

Enterium violaceum ATCC 12472 section 14 of 16 of the

Proteobacteria; Betaproteobacteria; Neisseriales; Pasteurellales; Chromobacterium.

1 GI:34104942

Enterium violaceum ATCC 12472

Enterium violaceum ATCC 12472

Proteobacteria; Betaproteobacteria; Neisseriales; Pasteurellales; Chromobacterium.

1 GI:303642

National Genome Project Consortium

Complete genome sequence of Chromobacterium violaceum reveals life and exploitable bacterial adaptability

1. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)

1 to 303642)

Los A.T.R., de Almeida, D.F., Almeida, F.C., de Almeida, L.G.P., de Almeida, R., Gonçalves, J.A.A., Andrade, E.M., R.V., Araújo, J., de Araújo, M.F.F., Filho, S.A., Azevedo, V., Bataus, L.A.M., Batista, J.S., Belo, S.A., Vander Berg, C., Bogo, M., Bonato, S., Bordignon, J., Brito, C.A., Burly, H.A., Camargo, A.A., Cardoso, D.D.P., N.P., Carraro, D.M., Carvalho, C.M.B., Cascardo, J.C.M., Chneir, L.M.O., Passa, T.B.C., Duran, N., Fagundes, N., L., Fancinetti, F., Farías, I.P., Felipe, M.S.S., L.P., Ferro, J.A., Ferro, M.I.T., Franco, G.R., N.N.S.A., Furlan, L.R., Gazzinelli, R.T., Gomes, E.A., F.M.M., Moreira, M.A.M., Nascimento, F.F., Niclaes, M.F., F.J.G., Oliveira, S.C., Palxao, R.F.C., Parente, V.A., F.O., Pena, S.J.D., Pereira, J.O., Petrich, M., S.R.C., Pinto, L.S., Porto, J.I.R., Pectrich, D.P., R.R., Reis, A.M.M., Rigol, L.O., Rondinelli, E., dos B.P., Santos, P.R., Schneider, M.P.C., Sena, H.N., M.R., da Silva, A.L.C., Silva, D.W., Silva, R., Simoes, I.C., Soares, C.M.A., Soares, R.B.A., Souza, E.M., Souza, K.R.L., C.C., Steffens, M.B.R., Steindler, M., Teixeira, S.R., T., Vettore, A., Wasseem, R., Zaha, A. and Simpson, A.J.G.

Submission

122-JAN-2003) Labinfo, LNCC - Laboratório Nacional de

Centro Científica, Rua Getúlio Vargas 333, Petrópolis, RJ

1, Brazil

Location/Qualifiers

1. 303642

organism="Chromobacterium violaceum ATCC 12472"

/mol\_type="genomic DNA"

/strain="ATCC 12472"

/db\_xref="taxon:243365"

complement(78..527)

/gene="rplI"

/locus\_tag="CV637"

complement(78..527)

/gene="rplI"

/locus\_tag="CV637"

/note="Identified by sequence similarity; putative; ORF located using Glimmer/Genemark/Blasx/COG0359"

/codon\_start=1

/transl\_table=11

/product="50S ribosomal protein L9"

/protein\_id="AAO61299.1"

/db\_xref="GI:34104942"

translation="MQITILEVAVNLQGLGVVNVNKGPRARNLFIQGRKKRATKATL

KEPDARRELEAKOASILLADAKVAREKLENEAVITITIAQAGVDGRLFGSVTNVDVAEAV  
TAAGVOIKRHEVLPNGPFRKAIIGYDIEIALHHDVATPIKIVVGEA"  
gene  
/gene="trpB"  
/locus\_tag="CV3638"  
/locus\_tag="CV3638"  
complement (549. .776)  
/gene="trpB"  
/locus\_tag="CV3638"  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/COG0238"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal protein S18"  
/protein\_id="AA061300.1"  
/db\_xref="GI:34104944"  
/translation="MARQLFRKKKCFRTAEGIKEDYKSVLDLKDFIAENGKIIPAR  
ITGKAVYQROLTPIAKRARFLAPYTDQH"  
complement (782. .1081)  
/gene="trpB"  
/locus\_tag="CV3639"  
complement (782. .1081)  
/gene="trpB"  
/locus\_tag="CV3639"  
/note="Identified by sequence similarity; putative; ORF  
located using GeneMark/Blastx/COG2965"  
/codon\_start=1  
/transl\_table=11  
/product="primosomal replication protein"  
/protein\_id="AA061301.1"  
/db\_xref="GI:34104945"  
/translation="MONRLVLTIVDREDALRYPAQLPVVEMWLKHOSRQSGFER  
DVACDIAVVGGEAKHAGKLAGRWNVTVTGLSQRSLRNPRLVNIETVEPVKG"  
complement (1100. .1474)  
/gene="trpB"  
/locus\_tag="CV3640"  
complement (1100. .1474)  
/gene="trpB"  
/locus\_tag="CV3640"  
/note="Identified by sequence similarity; putative; ORF  
located using Glimmer/Genemark/Blastx/COG0360"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal protein S6"  
/protein\_id="AA061302.1"  
/db\_xref="GI:34104946"  
/translation="MRHYEIVFIVHPDQSOVPAMIRYKGMVLAEGKIHREDWGR  
ROLAYPIQKLHKAHYVLMNVECOAETLAEIEHAFKFNDAVLRLTITMDRAISDSFM  
MDEKAKNLDAOPAEVEASA"  
complement (1631. .1927)  
/locus\_tag="CV3641"  
complement (1631. .1927)  
/locus\_tag="CV3641"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA061303.1"  
/db\_xref="GI:34104947"  
/translation="MKKLAWIALISGAHAAPIPETIYIKLHNGMLEAEIVSAGAP  
DYGSDNQGLASGAVTTRQLSMLANCGTPTVTVLTRDGVVVDIKDKKL"  
complement (1984. .2307)  
/locus\_tag="CV3642"  
complement (1984. .2307)  
/locus\_tag="CV3642"  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/COG1937"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA061304.1"  
/db\_xref="GI:34104948"  
/translation="MDDTHSSOACCHBEAOERSVVOGKAAALKRLRIEGOVGIGS  
GMESDRVCVDVLTQVAAVSALDAVALQGLENNMKCVARALQGGATGMATVELLIV  
VKYVR"

gene  
complement (2356. .3852)  
/locus\_tag="CV3643"  
complement (2356. .3852)  
/locus\_tag="CV3643"  
/note="Identified by sequence similarity; putative; ORF  
located using Glimmer/Blastx/COG2132/TC:9.A.10.1.1"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA061305.1"  
/db\_xref="GI:34104949"  
/translation="MEHAMORROMRWGAALALUSPMLRPAASAAHAGHGMAGQA  
APLPAALGLAGLQJLPDLKRLPLEKDGALRGCLPAALVSLPLFGKPAEFMAVY  
DSVPGRATLPEFGTGLRFDNRLPOPTTHMGLPIPSDDQGNPDHPVPGQGRDYR  
FSLPBDASGYWYHPRPHGTDAQAYGLAGVVPYKRNPDALPRMIVLSDLKD  
PDGKLIANSADNRDREGQFLVANGMOPRLALGRBSRQWRIMNTSARVLTALP  
AHEWMLVGTGGLEKPRVRYLLPRGAARLALVTRFPAKGVAGLVALPYAKGM  
GPEGQALPLDIEAGGMAVASLPERLRIDPLKASVTRRVFSDMPPAMFLI  
NGKTFDMRMDFTGOVGAVEWEIENDQADMDHPEHLHGTQFVTERHDKGMRAESYL  
AMRDVNVVPPGQVRLRFRODMGPRMFHCHIEHEDAGMMGTLDVRA"  
complement (3982. .4773)  
/gene="subB"  
/locus\_tag="CV3644"  
complement (3982. .4773)  
/gene="subB"  
/locus\_tag="CV3644"  
/note="Identified by sequence similarity; putative; ORF  
located using Glimmer/Genemark/Blastx/COG0483"  
/codon\_start=1  
/transl\_table=11  
/product="myo-inositol-1(or 4)-monophosphatase"  
/protein\_id="AA061306.1"  
/db\_xref="GI:34104950"  
/translation="MRPMLVAVKARRAGSVTORASLNDALIVEKKGNDPTEVD  
RAABEALINVILAEYPMGAILABESSGAKGVSSSEYEIIPBIDDTNPLHGHPOYGS  
IAMAHKQIOQAVVYDPRNDLFTASRGVAFCDLRASIKRPMNECLIGTFPVPS  
DQSLDQYGLMDVIAKTAGVREGAASLDLCVAGRTGEPFELNLKPMIDIAAGSL  
INOAGSVTPDPTEGQCFESGDI VAGNPKVLAQLDOLLAPHVAK"  
4946. .5680  
/locus\_tag="CV3645"  
4946. .5680  
/locus\_tag="CV3645"  
/EC\_number="2.1.1.-"  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/COG0565"  
/codon\_start=1  
/transl\_table=11  
/product="probable RNA methyltransferase"  
/protein\_id="AA061307.1"  
/db\_xref="GI:34104951"  
/translation="MQNIRVVLARPVHPNIGSARAMKTGSLRYLVEKVPPEE  
ANALASGAVDLLENATVGSRLREALSDVTVAACALTSRRELTTPLSPROTPELIAR  
ARDQEVALVFNGETFGLSIEVQCURLVTIPNPDYFSLNLMAVQVMVLEFSHT"  
GVDEYVRAEGEATQSEVDGMCGLQOAMDEIGYFQRNRSERILMRRTWFRHAGML  
REERDILRGFYKQVQRTASGRSDAD"  
5736. .6581  
/locus\_tag="CV3646"  
5736. .6581  
gene  
CDS  
Query Match 16.2%; Score 157.8; DB 1; Length 303642;  
Best Local Similarity 53.1%; Pred. No. 9.9e-13;  
Matches 451; Conservative 0; Mismatches 362; Indels 36; Gaps 4;  
Qy 22 CTCCTGCGCTGTGACGGGAAAGCCGCCCTCCCTGCGGGGGAGAGACCTT 81  
Db 66940 CTGGTGCGCTGGCAACGGCGCATGCGCCACGACCTGCCCTGGCAGTGAAGACCG 66881  
Qy 82 TACGCGCTCGGCTCTCCGAGGCTCTCTGACAGAACCGGGGTGAGACGCTCCCTCCC 141  
Db 66880 TATCGGGTGTGGCTGTAGAGATCATGCTGAGAGACCAAGGAGAGAGCGTCTCGAC 66821  
Qy 142 TATTACCGCGCTTTCGAGCGCTTCCACCTGAAGCGCCCTGCGGGCGGCTTCCCTG 201

Db 66820 TATTACCCGCTTCTGCGCCGCTTCCCGGACGTGGCCAGCTTGGCCGCCGCCGCGGTG 66761  
QY 202 GAAGAGGCTCTTAAGGCTGTGGCAGGGGGCGGGCTACTACCGCGGGGGGAAACACTTCAC 261  
Db 66760 GACGACGTGTGGCGCACTGGAGCGGCTGGGCTACTACGACCGGGGGCGCACTCTGAC 66701  
QY 262 CGCCTGGCGCAAGCTGTGAGGA-----GCTTCCCCGAGCTTCCGAG 306  
Db 66700 AAGCGGGGAAAGATGTGATGATGCTTTTGGCGGCGCTTTTCCCCCGAGCGGAGCGG 66641  
QY 307 CTTCGGGGGGCTTCTGTGTCTGGGCTTACACCGCGGGCGGGTGGGCTTCATGCGCTTC 366  
Db 66640 CTGAGGAGCAGTGGCGCGGTGGGCGGTCCACCGCCCGCGAATTTCCGCTTGCCTTC 66581  
QY 367 GGGAGCGGGTGGCGGGCGGTGAGCGGAACTTCGGAGGGTCTTCGCGCTTCCTTG-- 424  
Db 66580 GGGCGGGCGGAGCACTCTGGAGCGGCAATGTCAAAAGGGGTGTGGCGCGCTGCTTGCGC 66521  
QY 425 -----CCGGGAAAGCCCGCAAGGAGAGAGCTTTCCGCCCTCGCCAGGCG 471  
Db 66520 ATAGACGGCTTTCCTCGCGACAGGCGCATAGAGAAAGCATGTGGGCGCTGGCGAGAG 66461  
QY 472 CTCTCTC---CCGAGGCGGTGAGCCCGGGGGTGTGAACGAGCGCTCATGAGCTCGG 528  
Db 66460 ATCTGCGCGGCGCGCGCGCGGACATCGGCCCTATGTGAGAGGTGTGATGACTGGGG 66401  
QY 529 GCCACGGTCTGCGCTGCGGAAACGGCCCCCTTTGGCGGGCTCGCCCTTGAAGGCTTCTG 588  
Db 66400 GCGACGGTGTGACAGCCCGGCAAGCCGCGCTGTGACCGCTGTCCGATGTGAGACGGCTG 66341  
QY 589 CGGGGGAAGAGGCGCCCGGCGCTACCCCGCGCGTGAACCGCCCGAGAGCGCGGCGAAGAGAG 648  
Db 66340 GTGGCGGCGCGGGAGGGCGCGACCGCGAATCTGCCACGCCCGCGCAAAAAGGCTGTG 66281  
QY 649 CGCCTGTGCGCCCTCTCTCTCTCGGGCGGAGAGGGGTGCACTTGAAAAGCTTGAAGG 708  
Db 66280 CCATCTCGGACACCGGGATGTGTGGCCCGCATGAAMAAAGGTATGTGAGCGCG 66221  
QY 709 CGCTTCCAG---GAGCTCTACGGGCGTCCCGCTTTTCCCTGAGGAGCTTCCCGGCGG 765  
Db 66220 CGGCGCGCGACCGGCATCTGGGGCGGGCTGTTCGCTGCGGAGTTCGCCACACACTTG 66161  
QY 766 GAGCGCGCTTCCGGGTGAGGTGTAGGCCCTTGAAGGAGTGGCCAGCGCCCTCACCCAC 825  
Db 66160 GAATATGAAAGACTGTGAGCGCGCGGGGAGTGGCAGCATGTGCGCGCTGGCGCGAG 66101  
QY 826 CGGAGGCTT 834  
Db 66100 CTGAGCAT 66092

RESULT 6  
AL646059 190050 bp DNA linear BCT 02-SEP-2002  
LOCUS Ralstonia solanacearum GM11000 chromosome, complete sequence;  
DEFINITION segment 3/19.  
ACCESSION AL646059 AL646052  
VERSION AL646059.1 GI:17427391  
KEYWORDS  
SOURCE Ralstonia solanacearum  
ORGANISM Ralstonia solanacearum  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Burkholderiaceae; Ralstonia.  
REFERENCE  
AUTHORS Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,  
Arlat, M., Billault, A., Broctier, P., Camus, J. C., Catolico, L.,  
Chandler, M., Choisme, N., Claudel-Renard, C., Cunac, S., Demange, N.,  
Gaspin, C., Lavie, M., Moisan, A., Robert, C., Sautin, M., Schlex, T.,  
Sigstier, P., Thebaud, P., Whalen, M., Winkler, P., Levy, M.,  
Weisenbach, J. and Boucher, C. A.  
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum  
JOURNAL Nature 415 (6871), 497-502 (2002)  
MEDLINE 21681879

PUBMED 11823852  
REFERENCE 2 (bases 1 to 190050)  
AUTHORS Boucher, C. A.  
TITLE Direct Submission  
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CEPH, 27 rue Juliette Dohy, 75010 Paris, France, LMGH CNRS  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, and INRA  
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,  
F31326 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire  
INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.  
FEATURES  
source  
1..190050  
/organism="Ralstonia solanacearum"  
/mol\_type="genomic DNA"  
/strain="GM11000"  
/db\_xref="taxon:305"  
complement(163..750)  
/gene="RSC0382"  
/note="synonym: RS03350"  
complement(163..750)  
/gene="RSC0382"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PROBABILE RESPONSE REGULATOR TRANSCRIPTION  
REGULATOR PROTEIN"  
/protein\_id="CAD13910.1"  
/db\_xref="GI:17427392"  
/db\_xref="GOA:O8Y2F4"  
/db\_xref="SPTREMBL:O8Y2F4"  
/translation="MAPTRVLIVRDPLEBOVNLILIRAGLAAGFVDVGVNADLML  
PERIAQOPMVIVASESARADTIEHCYATGAPPIYLFPTNDASRIKAFPAIG  
TATVVDLKPTRKAVLDVAVARFERRARLRAQLDANKQLARKVLRAKGLLMQOM  
QLSDEAFKRLRKAMDRIKIVAAQFVIDVAG"  
937..1638  
/gene="RSC0383"  
/note="synonym: RS03351"  
937..1638  
/gene="RSC0383"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD13911.1"  
/db\_xref="GI:17427393"  
/db\_xref="GOA:O8Y2F3"  
/db\_xref="SPTREMBL:O8Y2F3"  
/translation="MTDFVVTTPAVTGILPVGAGGEFVRRVYCVGRNVAHAREMG  
SDPERPPFCKPDAVRVADGTGQFVYPTETDCHYEILVAIGTGGDIVES  
AAQHIYGAIGLDMTRDLQNAKKGGRWETGKAFDGSAPIGPIVPAQTVASPRGA  
ITLSVNDKAHQHGLSDLSIVPRTIAYLSRFLRPPGLITGPEGVGPISIDLM  
VGRIDGVELHVKVY"  
1642..2292  
/gene="RSC0384"  
/note="synonym: RS03352"



CDS  
1642..2292  
/gene="RSC0384"  
/EC\_number="5.2.1.2"  
/function="small molecule metabolism; degradation; amino acids"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PROBABLE MALEYLACETOACETATE ISOMERASE  
(GLUTATHIONE TRANSFERASE ZETA 1) PROTEIN"  
/protein\_id="CADI3912.1"  
/db\_xref="GI:17427394"  
/db\_xref="GOA:08Y2F2"  
/db\_xref="SPTREMBL:O8Y2F2"  
/translation="MSIKLYNPRSSASPRRIALIKGIPDYAPVHLKGEQAAD  
FVKLPBALPVLCDETVLNQSLAIVYEETHEPTLPGSADRAHTRALALATA  
CEIHLNPNRPVLKTKHTEFVNGDDARDWRHWRIGFPALETTLTQSPRTGACVAD  
TPTLADLCIVPOVFNGRFPVAVEDYPTLARVEHCMAOPAFORAPADPADASA"  
2289..2292  
/gene="RSC0385"  
/note="synonym: RS03353"  
2289..2927  
/gene="RSC0385"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PROBABLE TRANSMEMBRANE PROTEIN"  
/protein\_id="CADI3913.1"  
/db\_xref="GI:17427395"  
/db\_xref="GOA:08Y2F1"  
/db\_xref="SPTREMBL:O8Y2F1"  
/translation="MTGPGPSAPLIGALSFGSLIVATIGANAYVILROSLRE  
YVIGVUICALCDMALIALSVAGKGTISHPAMILTRVRNAGAAFLLAYGARARRA  
RGARLQARNKGRASHAQVLSALATLPHVLTDLTVLIGALGGRYAMPANAPAG  
GAMCASILWFSLEFGARLEPVPARPARVARDALIGAVMMALITLIMG"  
complement (2924..3910)  
/gene="ftay"  
/note="synonyms: RS03354, RSC0386"  
complement (2924..3910)  
/gene="ftay"  
/function="cell processes; cell division"  
/note="Product confidence : putative  
Gene name confidence : putative  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PUTATIVE CELL DIVISION PROTEIN"  
/protein\_id="CADI3914.1"  
/db\_xref="GI:17427396"  
/db\_xref="GOA:08Y2F0"  
/db\_xref="SPTREMBL:O8Y2F0"  
/translation="MPDTPAPAPALDMQADDIETVTPPVVEQARKKMSRLRSGLSK  
TSSKLTLPVGVYDEALFEELRALIMADAGDATYLLDELARRKKAORIEYABEV  
KTLARDLLELHPLEKTVLDGRDQPTVIMAGVNGKGTISGLKHFQITGQSVL  
LAAGDTFRAAREQLVIMGQRNNVTVAAGSGPAAYIPAVVNAARARGIDTWAQDA  
LRLPTQLHMBELKRRVIGKMAATPHETLIVDNTGQNALAQYKAFDDALGLTG  
LIVTKDGTAKGLTIAIARQRPVYFVIGGEVDLQFSAEFADALLG"  
4124..5629  
/gene="RSC0387"

CDS  
4124..5629  
/gene="RSC0387"  
/EC\_number="3.4.99.-"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PUTATIVE ZINC PROTEASE PROTEIN"  
/protein\_id="CADI3915.1"  
/db\_xref="GI:17427397"  
/db\_xref="GOA:08Y2E9"  
/db\_xref="SPTREMBL:O8Y2E9"  
/translation="MRSEFQSVNSARTGRMAGITLVALLAGAGNAPAPMAAQ  
AALLTGAQVTSQADTHEYRLANGLRLIVEDHRAPTVAHWWVHAGSIDHNGTGV  
AHMLBHMFGTRVCGEFSRRYVALGRENMTTDDFTMYFOITKSHADMALE  
ADRNANQLTDKEFKPEMNWVKEERRRIDSARATYEQMLAVLFNAAPRYNTIG  
PSDIDTIVDDADDPYHMYVAPNNATVVGVDVNPDEVFROAORTYKLOPHALPRY  
AOBEPQGVKRIWVKAPENPYVLAIVKAPLEBDEKVDVPYALEVSLVDGYNDA  
RLPNLVKRGDEKGRGLADDVNAAGYDGMNRPISFLDGVPADGHTTAETEOALRAOI  
DRIARDGVTREALEKRYVAQVYAAOIVKRDVFGQGMETGMAEMTGLSMRDLRILEKI  
KSVTPAQVQVAKTYFTEDNLVVAITLPPIDPKPARKPVPGRREGGLR"  
5629..6972  
/gene="RSC0388"  
/note="synonym: RS03356"  
5629..6972  
/gene="RSC0388"  
/EC\_number="3.4.99.-"  
/function="miscellaneous; hypothetical/global homology"  
Query Match 14.8%; Score 144.2; DB 1; Length 190050;  
Best Local Similarity 57.0%; Pct. No. 7.4e-11;  
Matches 322; Conservative 0; Mismatches 213; Indels 30; Gaps 2;  
gene  
CDS  
36 CCGGGAACCGCCGCCCTCCCTGCGGGGGAGAGACCTTACCGCTCTGGT 95  
19001 CCGGCAACCGCCGCATCCTGCGCGTGGAGAACACCGGTGACCTATCGGTGGCT 19060  
Qy 96 CTCGAGCTCTTTCGACGACAGCCCGGTGAGACAGCCCTCCCTATTACCCCGCTT 155  
Db 19061 GTCGAGATCATCTCGACGACAGCAGGTGTGCGCGGTGCGCTACACGCCGCTT 19120  
Qy 156 TCTGAGCGCTTCCACCCCTGAAAGGCGCTGCGCGGCTTCCCTGGAAGAGTCTTAG 215  
Db 19121 CGTGAGCGGTTCCCGACCTGACGAGCACTGGCCGCGACCCGGAACACGATGGC 19180  
Qy 216 GGTCTGACAGGGGGCGGCTACTACCGCGGGCGGAAACCTTCACCGCCTGGCCGAAG 275  
Db 19181 CGCTGGGGCGGCTGGGCTACTACACCGGGCGGCAATCTGACCGCTGGCGAGAT 19240  
Qy 276 CGTGAG-----GAGCTTCCCGAGCTTGGCCGAGCTTCCGGGGCTTCC 320  
Db 19241 CGTGCTGCGAGCATGCGCGGTGCTTCCGCGGATCCGAGCGCTGCGCGCTGCC 19300  
Qy 321 TGGCTCGGGGCTTACACCGCGGGCGGCGGCTTCATGCGCTTGGGGAGCGGGTGGC 380  
Db 19301 GGGCATGGGGGCTCCACCGCGGGCGGCAITGCGCGCTTTTCTGACGCGTGGCGGCG 19360  
Qy 381 GGCCTGAGACGGGACCTCCGAGGGTCTCTCCCGCTTTCG-----C 425  
Db 19361 CATCTCGAGCGGACGCTCAAGCGCGTGTGGCGGCTTGGCGATCGACGCGCTTTC 19420  
Qy 426 CCGGGAAGCCCAAGAGAGAGCTTTTGGCCCTTGGCCAGGGCTCTCTCCCGGAGG 485  
Db 19421 CCGGGAACACGCGCTCAGAGAAACGATGTGCGCATGCGGAGACCGTGTGCGCCCGGC 19480  
Qy 486 CGTGACCCGGGGGTGGAACACGAGCGCTCATGAGCTCGGGGCACGGTCTGCTGCC 545

Db		19481	GAGCGGATCTACTGTACACGCAGGGCCTGTATGACCTGGAGCCACGTCCTACAGCG	19540
Oy		546	GAAACGGCCCCGTTGGCGGCCCTGC	570
Db		19541	CGGCAAGCCCGCATGCTTACACCGCG	19565
RESULT 7				
BX640422				
LOCUS				
DEFINITION		257701 bp	DNA	linear BCT 14-AUG-2003
ACCESSION		BX640422		
VERSION		12/12		
KEYWORDS		complete genome.		
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REMARK				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
gene				
CDS				

```

BX640422                257701 bp      DNA             linear   BCT 14-AUG-2003
Bordetella pertussis strain Tohama I, complete genome; segment
12/12
BX640422 BX470248
BX640422.1 GI:33564552
complete genome.
Bordetella pertussis Tohama I
Bordetella pertussis Tohama I
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Alcaligenaceae; Bordetella.
1
Parkhill,J., Sebahia,M., Preston,A., Murphy,L.D., Thomson,N.,
Harris,D.E., Holden,M.T.G., Churcher,C.R., Bentley,S.D.,
Mungall,K.L., Cerdano-Tarraga,A.M., Temple,L., James,K., Harris,B.,
Quell,M.A., Achtman,M., Atkin,R., Baker,S., Basham,D., Benson,N.,
Cherevach,I., Chillingworth,T., Collins,M., Cronin,A., Davis,P.,
Doggett,J., Felwell,T., Goble,A., Hamlin,N., Hauser,H.,
Holroyd,S., Jagers,K., Leather,S., Moulé,S., Norbezczak,H.,
O'Neill,S., Ormond,D., Price,C., Rabinowitch,E., Butler,S.,
Sanders,M., Saunders,D., Seeger,K., Sharp,S., Simmonds,M.,
Skeldon,J., Squares,R., Squires,S., Stevens,K., Unwin,L.,
Whitehead,S., Barrett,B.G. and Maskell,D.J.
Comparative analysis of the genome sequences of Bordetella
pertussis, Bordetella parapertussis and Bordetella bronchiseptica
Online Publication
Nat. Genet. DOI:10.1038/Ng1227-10.1038/Ng1227
2 (bases 1 to 257701)
Sebahia,M.
Direct Submission
Submitted (06-AUG-2003) Submitted on behalf of the Pathogen
Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, E-mail: ms5@sanger.ac.uk
Location/Qualifiers
1..257701
/organism="Bordetella pertussis Tohama I"
/mol_type="genomic DNA"
/strain="Tohama I"
/db_xref="taxon:257313"
76..1266
/gene="tuf"
/locus_tag="BP3611"
/note="synonym: tufa, tufB"
76..1266
/gene="tuf"
/locus_tag="BP3611"
/note="synonym: tufa, tufB"
76..1266
/gene="tag"
/locus_tag="BP3611"
/note="Similar to Escherichia coli elongation factor Tu
(EF-Tu) TufA or TufB (duplicate of tuFA) SW-EFNU ECOIL
(Po2990) (393 aa) fasta scores: E(): 1.1e-122, 81.47% id
in 394 aa, and to Burkholderia cepacia elongation factor
Tu Tuf SW-EFNU BURCE (P3167) (396 aa) fasta scores: E():
3.4e-139, 92.65% id in 395 aa. Identical to BP0007,
100.00% identity (100.00% ungapged) in 396 aa overlap"
/codon_start=1
/tranl_table=11
/product="elongation factor Tu"
/protein_id="CAE43869.1"
/db_xref="GI:33564553"
/transl_table="MAKGFERTKPHVNVGITIGHVDHKTTITAITTVLSNRKGGA
RGYDIDABEEARKGTTINTSHVEYETRTYHAVDCGHADYVKNMTGAAMDGA
LIVYSANDGPQTRREHLISROGVAPYIIVLNADWDADAELAEVEMVERLLSK
YDFEGDDTPYIKVSATKLALBGDGGEIGEOAILSLNOALDTYIPTEBAVDAFLMPVE
DVFSIGRGTAVVGRTERGVYKAGEIEITVGIKPTNKTCTGVEMFRRLLOOGAGDN
VTGLRGTKREDVERGOVLAKPSINPHFDIFRAYIIISKSGCGHTFPFNNGYSPQFY
FTTIDRGTKEDVIDPADKEWMLPGNVSMTYKLAFIMEBGLRFALREGRIYAGAYVA

```

```

misc_feature
103..1203
/gene="tuf"
/locus_tag="BP3611"
/note="HMPFfam hit to PF00009, Elongation factor Tu family"

misc_feature
130..153
/gene="tuf"
/locus_tag="BP3611"
/note="ScaneRegExp hit to PS00017, ATP/GTP-binding site motif A (P-loop)."

misc_feature
226..273
/gene="tuf"
/locus_tag="BP3611"
/note="ScaneRegExp hit to PS00301, GTP-binding elongation factors signature. Confirmed by InterPro EMOTIF pattern match."

gene
1330..1641
/gene="rpsJ"
/locus_tag="BP3612"
/note="synonym: ruse"
1330..1641
/gene="rpsJ"
/locus_tag="BP3612"
/note="Similar to Escherichia coli 30S ribosomal protein S10 RpsJ Sm:RS10.ECOLI (P02364) (103 aa) fasta scores: E(): 8.4e-28, 76.47% id in 102 aa, and to Neisseria meningitidis 30S ribosomal protein S10 NMA0133 TR:Q9JRM2 (EMBL:AL162752) (103 aa) fasta scores: E(): 1.2e-29, 79.61% id in 103 aa"
/codon_start=1
/transl_table=1
/product="30S ribosomal protein S10"
/protein_id="CAE43870.1"
/db_xref="GI:33564554"
/translation="MNQKIRIRLRKAFDYKLIDGSAEIVDTAKTGAIVRGPRPLPT RIRRVYLSPHVNKTSRDQFEIRTHORLMDIVPDTKTVDALMRLLPGVDVEIAL Q"

misc_feature
1342..1629
/gene="rpsJ"
/locus_tag="BP3612"
/note="HMPFfam hit to PF00338, Ribosomal protein S10P/S20e"

misc_feature
1414..1461
/gene="rpsJ"
/locus_tag="BP3612"
/note="ScaneRegExp hit to PS00361, Ribosomal protein S10 signature. Confirmed by InterPro EMOTIF pattern match."

gene
2044..2688
/gene="rplC"
/locus_tag="BP3613"
2044..2688
/gene="rplC"
/locus_tag="BP3613"
2044..2688
/gene="rplC"
/locus_tag="BP3613"
/locus_tag="BP3613"
/note="Similar to Escherichia coli 50S ribosomal protein L3 RplC Sw:RL3.ECOLI (P02386) (209 aa) fasta scores: E(): 8.6e-48, 87.77% id in 211 aa, and to Neisseria meningitidis 50S ribosomal protein L3 NMA0129 TR:Q9JRH6 (EMBL:AL162752) (214 aa) fasta scores: E(): 2.4e-51, 69.48% id in 213 aa"
/codon_start=1
/transl_table=1
/product="50S ribosomal protein L3"
/protein_id="CAE43871.1"
/db_xref="GI:33564555"
/translation="WGRKVGWTRIFTEDEGSIFVTLDVSNRNVTQVKSLESDGVAAYQTVEPRATETVVRPOAGHYAKAGAEAGSILKEFLDPAAPAAEFAAGATVAESVFE GQGVDTSTSGKGFACITGRNNGSGORASHGNSRHRVSGISGQADGGRIFPGGR GSHMDVTRIVQNLVYRVADAEKGLVYKVAVGHGVDIVPAVKAPAKKA"

misc_feature
2053..2652
/gene="rplC"
/locus_tag="BP3613"
/note="HMPFfam hit to PF00297, Ribosomal protein L3"

```

```
misc_feature 2341..2412
/gene="rp1c"
/locus="BP3613"
/note="ScanRegExp hit to P500474, Ribosomal protein L3
signature. Confirmed by InterPro EMOTIF pattern match."
2693..3310
/gene="rp1d"
/locus_tag="BP3614"
2693..3310
/gene="rp1d"
/locus_tag="BP3614"
/note="Similar to Escherichia coli 50S ribosomal protein
L4 RpL SW:RL4 ECOLI (P02388) (201 aa) fasta scores: E():
5.9e-41, 58.33% id in 204 aa, and to Neisseria
meningitidis 50S ribosomal protein L4 NMA0128 TR:Q9JRA2
(EMBL:AL162752) (206 aa) fasta scores: E(): 2.4e-51,
65.02% id in 203 aa"
/codon_start=1
/transl_table=11
/product="50S ribosomal protein L4"
/protein_id="CAE34872.1"
/db_xref="GI:33564556"
/translation="MDLKLNDGGAATFSPADTIFGRDFNEALVHQIVAFQANARS
GNRAOKDAEYHSTKKRWROKGTGRABAGTSSPIARGGARAPNSPEENSOXVK
KNRYAGIRSLISQALREDRVAVDPTFTLESKTLAAKLSLSDSVLTITVDNEN
VYLAIRNDIPYAAVEPRYADPLSLIHYKVLITPRALQLEMLG"
2729..3301
/gene="rp1d"
/locus_tag="BP3614"
/note="HMMFam hit to PF00573, Ribosomal protein L4/L1
family"
3307..3603
/gene="rp1w"
/locus_tag="BP3615"
3307..3603
/gene="rp1w"
/locus_tag="BP3615"
/note="tag="BP3615"
/note="Similar to Escherichia coli 50S ribosomal protein
L23 RpL SW:RL23 ECOLI (P02424) (100 aa) fasta scores:
E(): 1.5e-14, 54.94% id in 91 aa, and to Neisseria
meningitidis 50S ribosomal protein L23 NMA0127 TR:Q9JX11
(EMBL:AL162752) (106 aa) fasta scores: E(): 6.9e-19,
63.83% id in 94 aa"
/codon_start=1
/transl_table=11
/product="50S ribosomal protein L23"
/protein_id="CAE34873.1"
/db_xref="GI:33564557"
/translation="MNAERLMQVILAPITVEKATFVAKRNOQVAFRVVADATPEIKR
AVELLFVQVESVQVILNRKGVKRFGRFVGRNRERKAYVALKQGEIDFAEVK"
3325..3600
/gene="rp1w"
/locus_tag="BP3615"
/note="HMMFam hit to PF00276, Ribosomal protein L23"
3604..4431
/gene="rp1B"
/locus_tag="BP3616"
3604..4431
/gene="rp1B"
/locus_tag="BP3616"
/note="Similar to Escherichia coli 50S ribosomal protein
L2 RpL SW:RL2 ECOLI (P02387) (272 aa) fasta scores: E():
2.4e-75, 67.03% id in 273 aa, and to Neisseria

Query Match 13.8% Score 135; DB 1; Length 257701;
Best Local Similarity 60.2%; Pred. No 1.1e-09;
Matches 250; Conservative 0; Mismatches 150; Indels 15; Gaps 1;

QY 36 CCGGAAAAACGCCGCCCTCCCTGCGGGGAGAGAACCCCTTACCGCGTCTGGT 95
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 217833 CCGGACGCGCGGCATCTGCGCGAGAACCCGCCATCCCTATCGCATCTGCT 217892
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 96 CTCGAGGTCTTCTGACAGACCCGGGTGAGACGAGCCTCCCTATTACCGCGCTT 155
```

```
Db 217893 CTCGGAATCATCTGACAGACAGCCAGTGGCCACGCTCATCTCCGTAATACACGCTT 217952
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 156 TCTGAGGCTTTTCCACCTGAAAGGCTTGGCGGCTTCCCTGGAAGAGTCTTAG 215
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 217953 CCGGAAAGTTCCTCCGATGTGGCCGCTGGCGGCGCGCCAGGAAGATGTATGCC 218012
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 216 GGTCTGCGAGGGGCGGCTCTACCGCGGGCGGAAACATCCACCG----- 263
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 218013 CTACTGGGCGGCGCTGGGCTACTAGCGGCGGCCCGCACTGACCCCTGCGCGCAGGA 218072
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 264 ---CCTGGCCGAAGACGTGAGAGAGTCTCCCGAGTTCGCCGAGCTTCGGGGGCTTCC 320
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 218073 AATCATGAGCGCTGCGTGGCGGCTTCCCGCGGCCGAAGATGCCACCTTCC 218132
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 321 TGTCTCGGCGCTTACACCGCGGCGCGGTGCGCTTCATCGCTTGGGGAAGCGGTGCG 380
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 218133 CGGCATGCGCGCTTCAACGCGCGGCGCATCGCGGCTTCCCTACGCGGAGCGCAGCC 218192
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 381 GCGCGTGAACGGAACGTCCGAGGCTCTCCCGCTTTCGCGCGGGAAGC 435
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 218193 CATCATGACCGGCAACGTCAAGCGGCTTCAACGCGCATCTCGCATCGAGGCG 218247
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
BX640435 346259 bp DNA linear BCT 14-AUG-2003
LOCUS Bordetella parapertussis strain 12822, complete genome; segment
13/14.
ACCESSION BX640435 BX470249
VERSION BX640435.1 GI:33574489
KEYWORDS complete genome.
SOURCE Bordetella parapertussis
ORGANISM Bordetella parapertussis
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Alcaligenaceae; Bordetella.
REFERENCE
AUTHORS Parkhill,J., Sebatina,M., Preston,A., Murphy,L.D., Thomson,N.,
Harris,D.E., Holden,M.T.G., Churcher,C.R., Bentley,S.D.,
Mungall,K.L., Cerdano-Tarraga,A.M., Temple,L., James,K., Harris,B.,
Quail,M.A., Achtman,M., Atkin,R., Baker,S., Basham,D., Basson,N.,
Cherveyach,I., Chillingworth,T., Collins,M., Cronin,A., Davis,P.,
Doggert,J., Felwell,T., Goble,A., Hamlin,N., Hauser,H.,
Holroyd,S., Jagers,K., Leather,S., Moule,S., Nordberg,K.H.,
O'Neill,S., Ormond,D., Price,C., Rabinowitch,E., Rutter,S.,
Sanders,M., Saunders,D., Seeger,K., Sharp,S., Simmonds,M.,
Skellon,J., Squares,R., Squares,S., Stevens,K., Unwin,L.,
Whitehead,S., Barrett,B.G. and Maskell,D.J.
Comparative analysis of the genome sequences of Bordetella
pertussis, Bordetella parapertussis and Bordetella bronchiseptica
JOURNAL Online Publication
REMARK Nat. Genet. DOI:10.1038/Ng1227-10.1038/Ng1227
REFERENCE 2 (bases 1 to 346259)
AUTHORS Sebatina,M.
TITLE Direct Submission
FEATURES
SOURCE
1..346259
/organism="Bordetella parapertussis"
/mol_type="genomic DNA"
/strain="12822"
/db_xref="taxon:519"
/complement(293..937)
/gene="ureG"
/locus_tag="BPP3852"
/complement(293..937)
/gene="ureG"
/locus_tag="BPP3852"
/note="Ortholog of Bordetella pertussis (BX470248) BPP166"
/codon_start=1
/transl_table=11
```

/product="urease accessory protein"  
/protein\_id="CAE39135.1"  
/db\_xref="GI:33574490"  
/translation="MHDISLSTTRTKTLPLRVGVGVSGKTTLLMEVCKAMYPQF  
DLIATNDIVYKEDRLTLTGALPERILIGETGCGPHTAREDAISINIIADOME  
QFPDDIVYFESGDNLAATPSPELSDLTLYIVASGEKIPRGGGGIRKSDLTFIN  
KTDLAPYAGADLAWEADTRMRDCKPFVWNCNLTGCGLDVIAFLTEGLFRG"  
complement(509..892)  
/gene="ureg"  
/locus\_tag="BPP3852"  
/note="HMPfam hit to PF01495, HyB/Ureg  
nucleotide-binding domain, score 4.5e-73"  
complement(848..871)  
/gene="ureg"  
/locus\_tag="BPP3852"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)."  
complement(962..1624)  
/gene="uref"  
/locus\_tag="BPP3853"  
complement(962..1624)  
/gene="uref"  
/locus\_tag="BPP3853"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3167"  
/codon\_start=1  
/transl\_table=11  
/product="urease accessory protein"  
/protein\_id="CAE39136.1"  
/db\_xref="GI:33574491"  
/translation="MHLSPPALPIGGSYQGLEAIEGLVHDEASTLAMESQLV  
VMKAEAPLWCLFEARAGDDAAHGWNQWFMASERRELRQETOMERSIARLOE  
LGWGTATRAVALRAPTLPAVHACACMMALPRAGAGAYVFSWLENOVAAIKGV  
PLQMGQRMLERLRAGLPVALDARARAGATPPRLDTFAPQVALSARHEQFSLF  
RS"  
complement(1094..1546)  
/gene="uref"  
/locus\_tag="BPP3853"  
/note="HMPfam hit to PF01730, Uref, score 9.7e-74"  
complement(1647..2264)  
/gene="ureE"  
/locus\_tag="BPP3854"  
complement(1647..2264)  
/gene="ureE"  
/locus\_tag="BPP3854"  
/note="Ortholog of Bordetella pertussis (BX470248)  
BP3167A"  
/codon\_start=1  
/transl\_table=11  
/product="urease accessory protein"  
/protein\_id="CAE39137.1"  
/db\_xref="GI:33574492"  
/translation="MKIANTFKRGAAGLDLSQAPGVTLTAERRRSRORLLDEGR  
GELGVAIERGQTLRDGDVLAEDGTVVVRALAEVATATPMOLARAHYHGNRH  
VLEIETAEQHLQFEDAVLIDMLAQGVVAMLELRAVFPEDVGAVGGHGHGDESGD  
DYALQAAVHAHEAHPHASHAGHGHVHSGHGHGKHGHEHDS"  
complement(1944..2249)  
/gene="ureg"  
/locus\_tag="BPP3854"  
/note="HMPfam hit to PF02814, Ureg urease accessory  
protein, score 1.7e-45"  
complement(2267..3982)  
/gene="urec"  
/locus\_tag="BPP3855"  
complement(2267..3982)  
/gene="urec"  
/locus\_tag="BPP3855"  
/EC\_number="3.5.1.5"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3168"  
/codon\_start=1  
/transl\_table=11  
/product="urease alpha subunit"  
/protein\_id="CAE39138.1"  
/db\_xref="GI:33574493"  
/translation="WTRISRSAAVAAIYGPVVGVDGRVRLADTLLEAVEKHHTING

misc\_feature  
complement(2273..3580)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="HMPfam hit to PF02802, Urease alpha-subunit,  
catalytic domain, score 0"  
complement(2969..3019)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="PS00145 Urease active site."  
complement(3548..3589)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="PS01120 Urease nickel ligands signature."  
complement(3584..3979)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="HMPfam hit to PF00449, Urease alpha-subunit,  
N-terminal domain, score 4.8e-83"  
complement(3979..4287)  
/gene="ureB"  
/locus\_tag="BPP3856"  
complement(3979..4287)  
/gene="ureB"  
/locus\_tag="BPP3856"  
/EC\_number="3.5.1.5"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3169"  
/codon\_start=1  
/transl\_table=11  
/product="urease beta subunit"  
/protein\_id="CAE39139.1"  
/db\_xref="GI:33574494"  
/translation="MIPGEITERGOIELVNGBPTLTIAVNEGDSPIOVGSYHRAE  
ANNAIVFDRELATGYRLNIPAGNAVFEFGMRRTVBLVAVGERRIFFGQKMGALK  
"  
complement(3985..4284)  
/gene="ureB"  
/locus\_tag="BPP3856"  
/note="This hit extended beyond the end of the feature by  
1 aa and was clipped.  
HMPfam hit to PF00699, Urease beta subunit, score  
3.2e-60"  
complement(4304..4921)  
/gene="ureJ"  
/locus\_tag="BPP3857"  
complement(4304..4921)  
/gene="ureJ"  
/locus\_tag="BPP3857"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3170"  
/codon\_start=1  
/transl\_table=11  
/product="urease accessory protein"  
/protein\_id="CAE39140.1"  
/db\_xref="GI:33574495"  
/translation="MKAMSKRAMLGSSGAALMLFSGAALAHPIGHILPGSMFPAAG  
WHPITGFDHLMLAVGMMSALTHTTAQAVLPMPLALLFAGAMMGAGVRLPAVE  
PYIWSLVLGLVLSRAVQGMAGPALVGFALPHGLAHGIELPGSGAAGFVAGFM  
LATLGLHAGLPAGPRLKXIMNMLSRALGVGIAGYALLFVAGAV"  
complement(4844..4921)  
/gene="ureJ"  
/locus\_tag="BPP3857"  
/note="Signal peptide predicted for BPP3857 by SignalP 2.0  
HMM (Signal peptide probability 1.000) with cleavage site  
probability 1.000 between residues 26 and 27;

misc\_feature  
complement(2273..3580)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="HMPfam hit to PF02802, Urease alpha-subunit,  
catalytic domain, score 0"  
complement(2969..3019)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="PS00145 Urease active site."  
complement(3548..3589)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="PS01120 Urease nickel ligands signature."  
complement(3584..3979)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="HMPfam hit to PF00449, Urease alpha-subunit,  
N-terminal domain, score 4.8e-83"  
complement(3979..4287)  
/gene="ureB"  
/locus\_tag="BPP3856"  
complement(3979..4287)  
/gene="ureB"  
/locus\_tag="BPP3856"  
/EC\_number="3.5.1.5"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3169"  
/codon\_start=1  
/transl\_table=11  
/product="urease beta subunit"  
/protein\_id="CAE39139.1"  
/db\_xref="GI:33574494"  
/translation="MIPGEITERGOIELVNGBPTLTIAVNEGDSPIOVGSYHRAE  
ANNAIVFDRELATGYRLNIPAGNAVFEFGMRRTVBLVAVGERRIFFGQKMGALK  
"  
complement(3985..4284)  
/gene="ureB"  
/locus\_tag="BPP3856"  
/note="This hit extended beyond the end of the feature by  
1 aa and was clipped.  
HMPfam hit to PF00699, Urease beta subunit, score  
3.2e-60"  
complement(4304..4921)  
/gene="ureJ"  
/locus\_tag="BPP3857"  
complement(4304..4921)  
/gene="ureJ"  
/locus\_tag="BPP3857"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3170"  
/codon\_start=1  
/transl\_table=11  
/product="urease accessory protein"  
/protein\_id="CAE39140.1"  
/db\_xref="GI:33574495"  
/translation="MKAMSKRAMLGSSGAALMLFSGAALAHPIGHILPGSMFPAAG  
WHPITGFDHLMLAVGMMSALTHTTAQAVLPMPLALLFAGAMMGAGVRLPAVE  
PYIWSLVLGLVLSRAVQGMAGPALVGFALPHGLAHGIELPGSGAAGFVAGFM  
LATLGLHAGLPAGPRLKXIMNMLSRALGVGIAGYALLFVAGAV"  
complement(4844..4921)  
/gene="ureJ"  
/locus\_tag="BPP3857"  
/note="Signal peptide predicted for BPP3857 by SignalP 2.0  
HMM (Signal peptide probability 1.000) with cleavage site  
probability 1.000 between residues 26 and 27;



OEASGKQMKVGFVSGSYISLMPQLLSAERKDYPRVDIVIBESTVVDLIRRLDHT  
LDVALYRPFVLEBETARTILLQADHLLVAPAGRAORDVALDELADPEFGYSRT  
HVPQMALIMYAFQVGVPHIAQEAIOVQITLISLVESGLAIVPKVACROAGSGVR  
LVNVPQALETIKVGIALVHPDNATPTANFVMACRLMOTEBPAGAR"  
2484. .2663  
/locus\_tag="BB4253"  
/note="HMMpfam hit to PF00126, DE Bacterial regulatory  
helix-turn-helix protein, lyxR family, score 7.5e-23"  
2526. .2618  
/locus\_tag="BB4253"  
/note="PS00044 Bacterial regulatory proteins, lyxR family  
signature."  
2733. .3356  
/locus\_tag="BB4253"  
/note="HMMpfam hit to PF03466, DE LyxR substrate binding  
domain, score 3.1e-43"  
complement (3406. .4377)  
/locus\_tag="BB4254"  
complement (3406. .4377)  
/locus\_tag="BB4254"  
/note="ortholog of Bordetella pertussis (BX470248) BP0418"  
/codon\_start=1  
/transl\_table=11  
/product="putative exported protein"  
/protein\_id="CAB34618.1"  
/db\_xref="GI:33577353"  
/translation="MHSRRLALALLSVALAOGSGAASADYPAHPKTIIVLPGSG  
ADTARFLSKHLAERKRPVENVNRCGSGFIAQAVATAPEDGTLTFASPSMTIN  
AAVPRUPYPAVDCPAPVAPLAFAPFPAVPAVPSRVSVAALAAARAPGQLNAGS  
TATYQVLELFEHQNGIKATHVYKGTSAADVAGVQVSVADVASALPIRGKSL  
RPLAVSTRIRIKDLPVPTMGESNGFPAVMTAFPAKVDPAIVASDAVALV  
RSQGRKAFMEQLGSEPFVSGPDLAFCORDELASMRIRAKLANIQE"  
complement (4306. .4377)  
/locus\_tag="BB4254"  
/note="Signal peptide predicted for BB4254 by SignalP 2.0  
HMM (signal peptide probability 1.000) with cleavage site  
probability 0.873 between residues 24 and 25;  
signal-peptide site"  
complement (3418. .4236)  
/locus\_tag="BB4254"  
/note="HMMpfam hit to PF03401, DE Uncharacterized protein  
family, UPF0065, score 4.4e-85"  
complement (4444. .5229)  
/locus\_tag="BB4255"  
complement (4444. .5229)  
/locus\_tag="BB4255"  
/BC\_number="4.2.1.17"  
/note="ortholog of Bordetella pertussis (BX470248) BP0417"  
/codon\_start=1  
/transl\_table=11  
/product="probable enoyl-CoA hydratase"  
/protein\_id="CAB34619.1"  
/db\_xref="GI:33577354"  
/translation="MTDFLRHRHGCVLLTMDPRTPNALSDAVALVQACARTG  
ADLSRAVLTGSHGVFSGNLTSLTVAGLSECPYHSRQATRSGLQRIPLVILN  
EPTTAAYNGAIGACDLCACDRLARQASFNESVVKGLTPGDGAMLLPRIVG  
MSRCELAFTGRSIDAARARELIVSEVFDALLPALALAOEIAHSGHALRTKX  
LIREQHTRLDTLLEISAGFOLAHHTOEHGALDAHSRRKG"  
complement (4660. .5190)  
/locus\_tag="BB4255"  
/note="HMMpfam hit to PF00378, DE Enoyl-CoA  
hydratase/isomerase family, score 3.9e-50"  
complement (5222. .6454)  
/locus\_tag="BB4256"  
complement (5222. .6454)  
/locus\_tag="BB4256"  
/note="ortholog of Bordetella pertussis (BX470248) BP0416"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAB34620.1"  
/db\_xref="GI:33577355"  
/translation="MSGPLSEGRILIDVTTVMGPYANQIIGDLGADVIVKPEPPAGNV

RGIGRHHADMGIGIFLHANSKRSIALDLKNPAGREALKVAATCDVLVYNVPEQAMA  
RLGISVAEVAOANPSILVYGMVGDORFVYARAAYDLDIGAGISLSVAGSDIP  
RYAPSAVDRIYVGIISAANAATAGLYNPARTOGOGALIVPEMTANHILIGDHMGOTF  
EPRIGRPGVPRILNADRRPVATRRNGHLCVLLYADKMKHARELLGOGRDADRRFD  
IGTRQNHATAYRLVSDIASRTTBEMLAETADITAMPVHTASLIRDPHLEAIG  
FGMDHSEGAERFMVAFNRWSGTRPPORGRHAPRLGSHSVLVLEAGYDDEAIALVR  
NGAALAAARQLPEGTDDH"  
complement (5681. .6250)  
/locus\_tag="BB4256"  
/note="HMMpfam hit to PF02515, DE CATB/BAIF family, score  
3.6e-24"  
complement (6650. .7405)  
/locus\_tag="BB4257"  
complement (6650. .7405)  
/locus\_tag="BB4257"  
/note="ortholog of Bordetella pertussis (BX470248) BP0415"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="CAB34621.1"  
/db\_xref="GI:33577356"  
/translation="MDYLVSLAQQPAAAMALATLIMEVVLIGIDNLIIFISILTNKLP  
OHQTKARIGIGMALIRLALGLVATVLTQVPELRFHGSWMDVILVAGGLTV  
MKATKEIHNVDRPDDMFGRKATLIGFAAVALIMLDLVYSIDSIITAVGKTEHI

Query Match 13.8%; Score 135; DB 1; Length 346287;  
Best Local Similarity 60.2%; Pred. No. 1e-09;  
Matches 250; Conservative 0; Mismatches 150; Indels 15; Gaps 1;  
QY 36 CCGGAAACCGCCGCCCTCCCTCGCGGGGAGAAAGACCTTACCGCGCTT 95  
DB 219781 CCGGACGCGCGCCACGATCTGCCCTGCGAAGAACCCGCGACCTTATGCACTTGCGCT 219840  
QY 96 CTCGAGAGTCTTTCGACAGACAGCCGGGTGAGCAGCCCTCCCTATTACCGCGCTT 155  
DB 219841 CTCGGAATCATCTGCGAGAGAGCAGGTGCGACGATCTCCGATTACACAGCGCTT 219900  
QY 156 TCTGAGCGCTTTCGCCACCTGAAAGCGCTCGCGCGCTTCCCTGAAAGAGTCTCTAG 215  
DB 219901 CTTGAAAGCTTTTCCGATGTGCGCGCTTCCCGCGCGCGCGCGCGAGAAAGATGCTATGCC 219960  
QY 216 GGTCTGCGAGGGGCGGCTACTACCGCGGGGCGAAGACCTTCCACCG----- 263  
DB 219961 CTACTGGGCGCGCTGCGCTACACGCGGGGCGCGCACCTTCCACCGCGCAGGA 220020  
QY 264 ---CTTGCGCGAAGCTGGAAGAGCTTCCCGGAGCTTGGCGAGCTTGGGGGCTTCC 320  
DB 220021 AATCATGACAGCTGCGCGCGCGCTTCCCGCGCGCGCGCGAAGATGCGCACCTGCGC 220080  
QY 321 TGTCTCGGCGCTTACACCGCGCGGGGCTCATGCGCTTGGGGAGCGGGTGGC 380  
DB 220081 CGGATGCGCGCTTCCACCGCGCGCGCATCGCGCTTGGCTTACGCGAGCGAGCGCC 220140  
QY 381 GCGGTGAGCGGAACGTCGGAAGGCTCTTCCGCGCTTTCGCGCGCGGAAAGC 435  
DB 220141 CATCATGAGCGGACAGTCAAGCGGGTGTTCACGCGCGCATTCGCGATGAGAGGC 220195  
RESULT 10  
SC0939116 293050 bp DNA linear BCT 11-FEB-2003  
LOCUS SC0939116 Streptomyces coelicolor A3(2) complete genome; segment 13/29.  
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 13/29.  
ACCESSION AL0939116 AL0935636 AL0949573 AL0949628 AL0949645 AL0949661  
AL0949707 AL0949763 AL0949819 AL0949841 AL079345 AL133236 AL133252  
AL645882  
VERSION AL939116.1 GI:24413781  
KEYWORDS Streptomyces coelicolor A3(2)  
SOURCE Streptomyces coelicolor A3(2)  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,

Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.M., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorek, A., Woodward, J., Barrett, B.G., Parkhill, J., and Hopwood, D.A.  
Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)  
Nature 417 (6885), 141-147 (2002)

JOURNAL MEDLINE 21996410  
PUBMED 12000953  
2 (bases 1 to 293050)  
Bentley, S.D.  
Direct Submission  
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces Sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA. E-mail: sdb@sanger.ac.uk  
On or before Oct 26, 2002. This sequence version replaced  
gi:4539555, gi:4582361, gi:4585581, gi:4585828, gi:4678647,  
gi:4678625, gi:4678897, gi:4757084, gi:4803679, gi:4808332,  
gi:5123647, gi:6469451, gi:6491805.  
Location/Qualifiers  
1..293050  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
57..713  
/gene="SC03282"  
/note="synonym: SCE39.32"  
57..713  
/gene="SC03282"  
/note="SC03282", possible precorrin-8X methylmutase, len: 218 aa; similar to many e.g. SW:COBH\_PSEDE Precorrin-8X methylmutase from *Pseudomonas denitrificans* (210 aa) fasta scores; opt: 868, z-score: 1000.1, E(): 0, (64.2% identity in 204 aa overlap) and SW:COBH\_MYCTU precorrin-8X methylmutase from *Mycobacterium tuberculosis* (208 aa) fasta scores; opt: 852, z-score: 982.0, E(): 0, (64.1% identity in 206 aa overlap)."  
/codon\_start=1  
/transl\_table=1  
/product="putative precorrin-8X methylmutase"  
/protein\_id="CAB40340.1"  
/db\_xref="GI:4582393"  
/db\_xref="GOA:O9X8F5"  
/db\_xref="SPTRMBL:O9X8F5"  
/translation="MTPTDSDKNREYEEYKDGAAIYQSFATTIPAEADLSGLPADVGR VAVMMIACGVNVDLVRLDLSWSPGVASRARELRAGAPVLCVRNVASGVTRARLPADN DVVCTLSDDPAVEPLAELRIGTRSAALRLMRDLRDGAVVAAGNAPTALFRLEMLIGEG APRPAIVIGVGVGAESKEALAAHPSGVEHLVVRGRGSGALAAAVAVIASEEE "

gene  
complement(745..1512)  
/gene="SC03283"  
/note="synonym: SCE39.33c"  
complement(745..1512)  
/gene="SC03283"  
/note="SCE39.33c, possible precorrin-6X reductase, len: 235 aa; similar to many e.g. SW:COBK\_RHOSO precorrin-6X reductase from *Rhodococcus* sp. (248 aa) fasta scores; opt: 873, z-score: 978.5, E(): 0, (55.1% identity in 247 aa overlap) and SW:COBK\_MYCTU precorrin-6X reductase from *Mycobacterium tuberculosis* (244 aa) fasta scores; opt: 690, z-score: 775.5, E(): 0, (50.4% identity in 244 aa overlap)."  
/codon\_start=1  
/transl\_table=1  
/product="putative precorrin-6X reductase"  
/protein\_id="CAB40341.1"  
/db\_xref="GI:4582394"  
/db\_xref="GOA:O9X8F6"

CDS

/db\_xref="SPTRMBL:O9X8F6"  
/translation="MHVLLIGTTEARRLRLAEILAAAGPACGRVTTSLAGVSRPTTP GERVVGFCGABGLAALRRHAYLLVDATHPFGCTTRNAHAATAAGVGLVLRP GMRVVEDVDVHAGSLAEALALPALRRVFLTTRGVLAVFALADPMLPVRSVDAP EGRRPTEVLRVLPGLQERLRRLRRHRLDIVLTGSGGAATAPKLTAREGLPV VVVRPVPVEGVVRVABEPGAAARVVAERARRSPRRP"  
1683..3818  
/gene="SC03284"  
/note="synonyms: SCE15.01, SCE39.34"  
1683..3818  
/gene="SC03284"  
/note="SCE15.01, partial CDS, conserved hypothetical protein, len: >436aa; similar to C-terminal region of TR:CAA22732 (EMBL:AL035161) another conserved hypothetical protein from *Streptomyces coelicolor* (860 aa) fasta scores; opt: 922, z-score: 1011.5, E(): 0, (39.5% identity in 425 aa overlap)."  
SCE39.34, unknown, len: >307aa; similar to TR:E1370406 (EMBL:AL035161) conserved hypothetical protein from *Streptomyces coelicolor* (860 aa) fasta scores; opt: 254, z-score: 277.0, E(): 4e-08, (29.2% identity in 277 aa overlap) TR:086875 (EMBL:U02894) hypothetical protein from *Streptomyces lividans* (547 aa) fasta scores; opt: 226, z-score: 249.7, E(): 1.3e-06, (33.0% identity in 197 aa overlap)."  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="CAD5311.1"  
/db\_xref="GI:24413782"  
/db\_xref="GOA:O8CJX1"  
/db\_xref="SPTRMBL:O8CJX1"  
/translation="WRKARGAKAGCAVTCDPIDGAAYLSALFANSPOGLFVDSO KVRKYNPAGRGVRLPREDDVVGDIATVFGFEVVEVLQIDELVAGLGEALVRG RABSGPRTTAAVEVSLPRLDLDGKGLVGVSEVTERQALAEVRLVLSVHTVGS TLVDRTGDELVALVAFADATVDDLDGCVAAABGAPPEVPLRVAFAPPA EVSRKGGSDRPLPEPTPYTOVNDMRARLVSSDDPMLSDPALBPLVATGMSI VAPLRAIVYGLITLVRHRTDPEERDVLVARAATASVHLNANSYRHHVAAA LQRLQGLAIPLSAVETTAHYVLPESGQDMFDPVPSGRTVALVVDVGHGIEAA LMGQLRAVRLTARLDETDELNLHDEVASQLDASGPGSNTVATCAVTVAPVSR HCTVNRAGHAPVVIDDGVRAVDVQGPRLGTGGVHFVPAVLELAPGLLALHTN GLTRFGNDEGEAARLRLEHILASTTRPLEOELCDTAVVHMAFSRHDDVLLARTALP PEHVADWTLTPADASVGTARLVDRGLASGLDAAVTTGLVSELVETNAIRYGGPV RLRLIRGRGLTETVDANSGASPHLRARNDGCRGLFTIMRSTHMGVHSHDKT IWSQRJLDAGPPTSGLDLDFVTDAAGL"  
complement(3855..9173)  
/gene="SC03285"  
/note="synonym: SCE15.02c"  
complement(3855..9173)  
/gene="SC03285"  
/note="SCE15.02c, large glycine/alanine rich protein, len: 1772aa"  
/codon\_start=1  
/transl\_table=1  
/product="putative large glycine/alanine rich protein"  
/protein\_id="CAB41271.1"  
/db\_xref="GI:4678899"  
/db\_xref="SPTRMBL:O9X877"  
/translation="MATAPKXATGTHDTGDTADAGAVEVLRDLAAQADLARRAGL LMARVEPFSARLELASTERLTRHEGVPCDDLAADVADQADVEALVADAEATTA BAGVVARLDELADGLRTVDDVAVELDVGAATVATLLERVAALVGNRRARLTDAR RRLALDREGRAAEFTAFALGOAVTAAALADLPERDCDQRLARLALDELSEFAER DGFALBLADRTETIYDALARKKALSTPRARRABALASARIMETIARRCATLADAD AVSTYFASDPMKRVKVRTAEELRALGDSVAEELDGLKSKARGASALNRRTLVAD

gene  
CDS

DGRTLRGLAHNRPAVNTQPLDITLVDPDGLAFLTGTDYRSPVTPDPAATRAHWRT  
 LPSBPGVYRAHRLARLLRQHGASALAAADLPALVREAQEAVIDGYEGYVHDHA  
 TLVLTALLPLEYKAGTIVHBPAPAAQQLFWAGTTPAPAMTRRARSLAPATDTG  
 LSAATLALBEEELAGAVGTWPGVEVAGVAPGGAAGVAGAPABGRAGVAPATGAV  
 GSPASGSPSGAGAPLGAAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG  
 AGGTRSAEGFAG  
 LGAGTRTLEKPRPTVGTGPADEDLAALDLAAGQALAEANTSSYALAGGTGTPGDL  
 AEAVAELCEPLPRIDGVPPTATAGLGHPRITGGRLQRLRDEFLATPAPAAHD  
 VPERAVORRRRTALVAGARARLRUDHPRVMSAFVNRRLVDEVLPLVGLSLAKOG  
 ATGDKRTDGTGGLLLSPGKYKTTTIVEYAEKGLMLVVGSGPALGHGVTSLPAD  
 APNATARQEVKINFALASANTLLYLDIOHTSPBLOKFIPLCDATREVDGVMNKA  
 ERTYDLRGRFVACMAGNPTESGARFOVPMELADVNVLKIPALVETKEAASLST  
 ENALTJANVLAPLAGRDADELVLRLATDTPARARLDHAPVPALELITVLRHL  
 VAAREYLAANVAATIASASBDEFTTTPAPRLOGSYNNMKIAQRPVNVNDERAAV  
 LDHVTYEAQTLTGAEANLKLKLAELTLPEDQADMAEVRTAHWARTLGLGDDP  
 LTRAVAAAGLILAAVAEVAESITRAADPRHLNPHARHAAAGTTR"  
 complement (9259, .11292)  
 /gene="SC03286"  
 /note="synonym: SCE15.03c"  
 /complement (9259, .11292)  
 /gene="SC03286"  
 /note="SCE15.03, putative secreted protein, len: 677 aa;  
 weak similarity to SW:YQIK\_ECOLI hypothetical protein from  
 Escherichia coli (553 aa) (fasta scores: opt: 542, z-score:  
 501.2, E(): 1.3e-20, (27.0% identity in 504 aa overlap).  
 Contains possible N-terminal region signal peptide  
 sequence also contains a highly hydrophilic central part  
 of the protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative secreted protein"  
 /protein\_id="CAB41272.1"  
 /db\_xref="GI:4678900"  
 /db\_xref="SPTREMBL:O9X878"  
 /translation="MEVITGLAVCLLVTVLFGVSLFRVYEGKALIVSKMRVD  
 VTTPQVNVLPVHKAEMWDSVKTITETPAKEGGLCRNIRADRIISFVNVNTAE  
 DVKVAQVGTARASDRDTQLFLHAKFSBALKTGKQDLFDLIRREBELYRIIE  
 IGVLDLGHLEDAIDYLEQTLQDLPANVLDAQIRITETLVAENHTEARTAE  
 EKETIRQNVARAEVLELRQADAEIKOREIDTTRAEAEATVAEVEEELNOAA  
 FLTEREOLGVOREVAVAKNERVIAVNERVIAVNERVIAVNERVIAVNERVIA  
 AEKEVEERDIAEVIRAVAVDRVTAEGEBSIKKIRAVAEERBDQTVIIAAEQAO

gene  
 CDS  
 complement (9259, .11292)  
 /gene="SC03286"  
 /note="SCE15.03, putative secreted protein, len: 677 aa;  
 weak similarity to SW:YQIK\_ECOLI hypothetical protein from  
 Escherichia coli (553 aa) (fasta scores: opt: 542, z-score:  
 501.2, E(): 1.3e-20, (27.0% identity in 504 aa overlap).  
 Contains possible N-terminal region signal peptide  
 sequence also contains a highly hydrophilic central part  
 of the protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative secreted protein"  
 /protein\_id="CAB41272.1"  
 /db\_xref="GI:4678900"  
 /db\_xref="SPTREMBL:O9X878"  
 /translation="MEVITGLAVCLLVTVLFGVSLFRVYEGKALIVSKMRVD  
 VTTPQVNVLPVHKAEMWDSVKTITETPAKEGGLCRNIRADRIISFVNVNTAE  
 DVKVAQVGTARASDRDTQLFLHAKFSBALKTGKQDLFDLIRREBELYRIIE  
 IGVLDLGHLEDAIDYLEQTLQDLPANVLDAQIRITETLVAENHTEARTAE  
 EKETIRQNVARAEVLELRQADAEIKOREIDTTRAEAEATVAEVEEELNOAA  
 FLTEREOLGVOREVAVAKNERVIAVNERVIAVNERVIAVNERVIAVNERVIA  
 AEKEVEERDIAEVIRAVAVDRVTAEGEBSIKKIRAVAEERBDQTVIIAAEQAO

Query Match 13.6%; Score 132.2; DB 1; Length 293050;  
 Best Local Similarity 53.9%; Pred. No. 2.6e-09;  
 Matches 381; Conservative 0; Mismatches 293; Indels 33; Gaps 4;  
 20 CCTCTCGGCTGTACCGGAAAGCGCCGCCCTCCCTGCGCGGCGG---GAGAAAG 76  
 Db 89501 CCTCATCGACTGTTCAGACGACGCGCGCATCTGCGTGGCGCGCGCGCGCGCG 89560  
 QY 77 ACCCTTACCGCGTCTGTCTCGAGAGTCTTCTGACAGACGCGGAGTGAAGCGCC 136  
 Db 89561 GCGCCCTGGGTGTGATGTGACGAGATTATGTCTCAGACGACGCGGTCACCGTTC 89620  
 QY 137 TCCCTATTACCGCGCTTCTGAGAGCTTTCACCTGAAAGCCTTGGCGCGGCTT 196  
 Db 89621 AGCCCGTCTACGACGATGCTGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGAG 89680  
 QY 197 CCTGTGAAGAGTCTTGAAGGTCTGCAAGGCGCGGCTTACTACGCGCGCGGAAAC 256  
 Db 89681 CCGCCGCGAGGCGCGTGGCGCGCGCTGCGCTACCGCGCGCGCGCGCGCGCGCG 89740  
 QY 257 TCCACCG-----CTGGCCCGAAGCGTGAAGAGCTTCCCGGAGCTTTCG 301  
 Db 89741 TGCACGTCGCGCGCGCGCATACGCAAGCGGCGCGCGCGCGCGCGCGCGCG 89800  
 QY 302 CCGAGCTTGGGCGCTTCTGTGCTCGGCGCTTACACCGCGCGCGCGCGCTTCATCG 361  
 Db 89801 CCCAGCTGTGCGCTGCGCGCATCGGAGATACGCGCGCGCGGCTTCCTCTTCG 89860  
 QY 362 CTTTGGGAGCGGCTGCGCGGCTGAGACGCGGAGTCCGAGAGTCTCTCCGCGCTCT 421  
 Db 89861 CGTACGCGCGCGCGCGCGCGCTGAGACCAACGTCGCGCGGCTCTCCCGCGCGG 89920

QY 422 TCGCCCGGGAAGCCCGCAAGAGAGAGCTTTTCC-----CTCGCCAGG 469  
 Db 89921 TCACCGGTGTGACGATCCCGCGCAACGCGCACCGCGCGCGAGCGAAGCTGCGCGG 89980  
 QY 470 GCTCTCTCCCGAGGCGCGTGAACCCGCGG---TGTGAAACGAGCCCTCATGAGCTCG 526  
 Db 89981 CCTCGTGTCCGAGAGAGAGAGAGAGGCGCGCGCGCGCGCGCGCTCGATGAGACTCG 90040  
 QY 527 GGGCCACGCTTCTGCTGCGCGAAACGCGCCGTTGGGCGGCTGCCCCCTAGGCGCTTCT 586  
 Db 90041 GTGCGCTGTCTGACGCGCGAAAGAGAGTGTCTCACCGCTGCCATCGCGCGCGAGT 90100  
 QY 587 GCGCGGGAAGAGAGCGCGCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGAGAG 646  
 Db 90101 GCGCGTGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGCT 90160  
 QY 647 AGCGCTGTGCGCT 693  
 Db 90161 AGCGCGGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGCT 90207  
 RESULT 11  
 LOCUS AP003012/c 346547 bp DNA linear BCT 15-MAY-2001  
 DEFINITION Mesorhizobium loti DNA, complete genome, section 19/21.  
 ACCESSION AP003012 BA000012  
 VERSION AP003012.2 GI:14026998  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mesorhizobium loti  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Phyllobacteriaceae; Mesorhizobium.  
 1 (sites)  
 Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,  
 Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,  
 Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,  
 Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M.,  
 Takeuchi, C., Yamada, M. and Tabata, S.  
 Complete genome structure of the nitrogen-fixing symbiotic  
 bacterium Mesorhizobium loti  
 DNA Res. 7 (6), 331-338 (2000)  
 JOURNAL  
 PUBMED  
 MEDLINE  
 21082930  
 11214968  
 2 (bases 1 to 346547)  
 AUTHORS  
 Kaneko, T.  
 TITLE  
 Direct Submission  
 Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research  
 Institute, The First Laboratory for Plant Gene Research; Yana  
 1532-3, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: kaneko@kazusa.or.jp)  
 URL: http://www.kazusa.or.jp/rhizobase/  
 Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)  
 On May 11, 2001 this sequence version replaced gi:11994987.  
 COMMENT  
 FEATURES  
 source  
 1..346547  
 /organism="Mesorhizobium loti"  
 /mol\_type="genomic DNA"  
 /strain="WAF30309"  
 /db\_xref="taxon:381"  
 /complement (4, .507)  
 /gene="m117507"  
 /complement (4, .507)  
 /gene="m117507"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB53953.1"  
 /db\_xref="GI:14026998"  
 /translation="MPAGGPTMPMDVFWRAVAGTATVMDLAILLHKAFOGPR  
 NMGPGRVNMHGDVRFHDDIGEAEPYAHETALGVHVFVGIYGIILVLVAGTGLW  
 AAPITVPAFIIIGITVVGAWFLAPGKAGAAASKRPVPMIRALNIVSHTVFALGLW  
 GTALLIR"



gene  
complement (574. .1056)  
/gene="m117508"  
CDS  
complement (574. .1056)  
/gene="m117508"  
/codon\_start=1  
/transl\_table=1  
/product="peroxiredoxin-like protein"  
/protein\_id="BAB53954.1"  
/db\_xref="GI:14027000"  
/translation="MTISVGDKLPEATFKMTADGAKITSAEIPEKRVVLCFVPGCA  
FTPTCSNNHLPGYLENDLAILARGVDITAVSVNDVHVMGAMARFTGSEDKILTLAAG  
SADPAKAVGLDNDLSAGMGMLRSRFSMTVDDGKVTALNITREKGVDESAAHILGOL  
"  
gene  
complement (1114. .1950)  
/gene="m117509"  
CDS  
complement (1114. .1950)  
/gene="m117509"  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB53955.1"  
/db\_xref="GI:14027001"  
/translation="MOSLIKILLGALIGASVLPASASSAMVNSEGGKRVLYTSGKP  
DEAGRIQGVLDIALKPGMTYWRDPGAGVPPOLNVSATNIADATLSFPAPORHDDG  
VKGAGVNYVSLPVFTLSAKDSAVIDADVETLIGICETTCIPVOTRLTVDPSDDPN  
ADDAALYKASFPALPAPARSDPGINVLPGHETLYVARSFPGDEADDFYAGSERDTM  
FGAPRSREKDGKILFTVPLIDRPSATPTDGLHTLTLSAGAVEGLLPF"  
2122. .2730  
/gene="m1r7511"  
2122. .2730  
/gene="m1r7511"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB53956.1"  
/db\_xref="GI:14027002"  
/translation="MDLLRHKKTAAAGRGFLDDOFLIAMPKMDRFTRSVLYICASHD  
EGAMGLINOTQOMLPFDLYVQIGINMEQEIIRLPAQARDPVVRNGCPVDSRSPVUH  
SGDRVSSLSLVSDDICLTATVDILRAISSGRGRHALLMAGISGKAGQULETIAAN  
GWLTPASPELFPADIERYDRIIASIGLIALSLAAGA"  
complement (2752. .4416)  
/gene="m117513"  
complement (2752. .4416)  
/gene="m117513"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB53957.1"  
/db\_xref="GI:14027003"  
/translation="MAGSGDTWMDVLRDRVYTKPDVSIQGLAPNSLGGAAHMLP  
VLHADDDTFRITLDVLEHRRGVAQNFRLAGDGHYHMFALARPVIGSDGEVIRC  
VGVAVDTQEKSEERLHDVADHNLGFLPRELFPNRLBAIISIARTEEKVRFVIV  
IDIDRFKQVNDGLISAGDTILTITARRHLRLKXSLSPFADGOFALMLSEODAR  
RIGVADAIRHAINNPITPAKREIVLTASIGLITWTSQTSAEEDMVADELAMHQAAR  
FGGDRIEPRPARFTVCTDRIQESDILRAIRERETFLAYOPIYRLDGSYAGREALL  
RMWDRPRGMLPPADFIIVASCGLIYVQGLFAMQAAEDLAGOKQIGDAPLSVNL  
SSRQLIRDLVSDVRSIVARANLKPFCRLLETSLVMDNEQOTAHVLTKLQKQIGIL  
SLDFFGTGYSLATLRFPPFTIKIDKSFYVDNTPKRAVLKSNVMAHBEIGLSVAE  
GISDERDALELRQWGCCEVOSFMFGAMPDQVYKTLKEQYPLTQA"  
complement (5644. .6240)  
/gene="m117514"  
complement (5644. .6240)  
/gene="m117514"  
/codon\_start=1  
/transl\_table=1  
/product="ribosomal-protein-alanine N-acetyltransferase"  
/protein\_id="BAB53958.1"  
/db\_xref="GI:14027004"  
/translation="MFAIPFPRDLPALKGNLVTLRVPYTNVREWSLGRGSAFLE  
PWBRWTPDELDRTRAMRLISRYREDAQGTALAFIFEEKSGCLAGGILTNIRHGV  
SOSGVGWIGDERGGRGLMDAVKAVFAFDLRLHRTBAACIPONINSIRVLEKA  
GFREGLIRSLYLRNGIMODHYLVARLADPDLDACTD"  
complement (6256. .7548)

CDS  
/gene="m117516"  
complement (6256. .7548)  
/gene="m117516"  
/codon\_start=1  
/transl\_table=1  
/product="processing proteinaase"  
/protein\_id="BAB53959.1"  
/db\_xref="GI:14027005"  
/translation="MGYVSRLSNGLTVAETLPSITSVALGAMVKGARREDBEHG  
LADLHEMARKGTARTAFETASIEDVGEINATSEVETTSYARVLSDVPLAVDI  
LADLQSEEDPQDLEREQHVILQELGAADTPDILVDRFTEFAFPHQTGRSILGT  
PDYVSTSKQLHDFIERQYGAERMYVLAAGDILKHDFNREVERKQLGFSRKAOSTIP  
QYAGVYSGDFREDRLMDAQIVLGFEGRAVHRDVFYASOYLSMLTGGMSRLLFOEVR  
EKRGICYSYAFHMGFSDDGTGICGHATGOSDITAEIVPIIDELQKAGENTLOEILDR  
ARAQYRAGLIMSASPARASQIAROLLGRLPAKELMERLSALTVERLTDLSRM  
FSTKPYLTAVPGVTLAPYALIDSLSGTQTRARLAV"  
complement (7563. .8960)  
/gene="m117517"  
complement (7563. .8960)  
/gene="m117517"  
/codon\_start=1  
/transl\_table=1  
/product="threonine synthase"  
/protein\_id="BAB53960.1"  
/db\_xref="GI:14027006"  
/translation="MQYVSRGEAPRALGFSDDAVLAGIARDGLVYPREMPOFSSEIR  
AMGALPDLAIRYLPFLGSEIPAPYFEBVLVEAYATFFHEAYCPVLVGTNTFVLE  
LFHGPTLAFKDVAMQTLARLMDHVARLDHATTVGATSGDTGGAALDAFAGRSRTDI  
FIFLPHGRVSPVOQOMTTSKAEVHALATEGNDDCGILKDMFNDAFDRVALSG  
VNSINMARIMAOIYYFSSALSICGAPRPVSPFTVPTNGFDIFAGYAAKKMGLPIERL  
VIATNDNDILARTFATGEVYRTKGFATTSQSMOIQVSNFRLIFEASNDPAATVVR  
MDGLKOGAFITRESREINGRSEFDDARATMDVPAATIRSTLASNYLDPHTAAAH  
VAGKASGAVPMVVLGTAFHPAKFPAAVEAASVSPALPWLGGMLTBEKTYLPSDL  
KMVEDYVSRRAAR"  
8930. .9817  
/gene="m1r7518"  
8930. .9817  
/gene="m1r7518"  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB53961.1"  
/db\_xref="GI:14027007"  
/translation="MLPRATSHIACLSPPVAAMLSVHKVNYFCAGSSRIPAAALIVYT  
SFTRGKCSFMAFRTLSGRFLAGALTGFMILAAAGCSGDGILNLFGEKDPFAPPP  
QDPKILASQUNAYCPRVTYVDGTAFETVYANEVKAPKPKKTGDAADAAQAQAAAD  
AAATPPDSSARIITYOASIAVDRDSCSADQLAMKIVAGKITVGPFTPTGTTMPYR  
TAAWGHVDVLSQIHQYQVOTDPSVATQVFTDTSNVVPAPFTADQYVAYGDETAP  
KATTDKPKTKHRRKAATTN"  
complement (9860. .10549)  
/gene="m117519"  
complement (9860. .10549)  
/gene="m117519"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB53962.1"  
/db\_xref="GI:14027008"  
/translation="MPQPDVIFPCDDVLDVSEIIIAARVAEELITLAGYEISAEIIAE  
SYAGLTKPDILMRIEESKILPFQASLIDRAEELVDRSDVRAIEGVREAVSVTQ  
RCTCSNRSRRIRIEKXVLLPFPARIRISALEIPSKTKRPADVDLFAABRIGAMP  
ANTFVIEDSVGVAGARAAGRVIYFTGASHYIPGHADALTGCAETVIRRMALIKSV  
IALSEWSADA"  
10726. .11859  
/gene="m1r7520"  
10726. .11859  
/gene="m1r7520"  
/codon\_start=1  
/transl\_table=1  
/product="adenine DNA methyltransferase"  
/protein\_id="BAB53963.1"  
/db\_xref="GI:14027009"  
/translation="MSAVRLHLSHPQOSEWLDITLKQDCVAALDRLPKSVYVIF

ADPPYNQLDGDHLPDQSKVDADVHDWQFENFEAYAFETRAWLLAARVUKENGT  
WVIGSYHNI FRVAGMODLGFWILNDVYWRKTPNPNRGRFQAHETMTWASRDQK  
GKGYTFENYALKASNDLIQMRSDMLFPICTGSERLKNGDMLHPTQREPLAIRM  
ASTKRGDVLDPFGSGTTGAVAKRIGFVIEBOAYIDANRIRIDAVPLEDADL  
TVLTKRABEPRAVPSLIDTGLMTPGVTLYDKRMWAKVRADGVVALGDSAGSIHKI

Query Match 13.4%; Score 130.2; DB 1; Length 346547;  
Best Local Similarity 55.8%; Pred. No. 4.5e-09;  
Matches 324; Conservative 0; Mismatches 233; Indels 24; Gaps 3;

33 GATACGGGAAAAAGCCCCCCCCCTCCCTGGGGGGGGGAGAGACCCCTTACGGGCTCT 92  
13560 GTGGGGGGGAGACCCCGGAGACATGACGCGCGCTGGCCCGACCCCTATCGACTTG 13501  
93 GGTCTCCGAGGTCTCTTCTGACAGACCCGGGTGAGACGCGCTCCCTATTACCGCGC 152  
13500 GCTGTCCGAGATCATCTGACAGACAGCGGTGAGGGGGGTGAATCTTACTTCCGCC 13441  
153 CTTTCTGAGGCGCTTCCCAACCTGAGGCGCTGGCGCGCTTCCCTGGAAGAGTCT 212  
13440 CTTTGTGAGAAATGGCCGATGTGAGGCGCTGGCGCGCGCGCCGACGAGAGAGTCTAT 13381  
213 TAGGCTGAGGAGGGGGGGGCTACTACGGGGGGGGAACACTTCACGCGCTGGCGCG 272  
13380 GAGGCGCTGGCGGGGCTGGCTATTACTCCGCGCGCGCAACCTTACAGGCGCTGGCG 13321  
273 AAGCGTGAAGAGC-----TTCCCCGAGCTTGGCGAGCTTGGGGGCTTCC 320  
13320 TCTGTGGGGGGGGCGGTGGTGGCGCTTCCCGATACGAGGCGGATTTGAGAGACTGCG 13261  
321 TGGCTCGGAGCTTACACCGCGGGGGGGGCTTCATCGCTTGGGGGAGCGGTGGC 380  
13260 CGGCAATCGGCGCTTACACTTGGCGGGGCAACAGGGGATCGCTTGGACCGCGCGCGC 13201  
381 GCGGCTGAGCGGGAAGTCCGAGAGGCTCTCCCGCTCTTGGCGCGGGAAAAAGCCCA 440  
13200 CTTGTGATGCGCAATGTGAAAGCGCTCATATGAGGCTGTTTTCATACAGACGCACT 13141  
441 GG-----AGAGAGACTTTCGCTCGCCGCAAGGCGCTTCCCGAGGGCGTGA 491  
13140 CAAGCAAGCAAGACGAGATACGAGCTTGTGAGGCGCATGTGCTCCGGCGCAAG-- 13083  
492 CCGGGGGGTGTGGAACGAGGCTCATGAGCTCGGGGGCGACAGTGTGCTGCGCAAGG 551  
13082 -CCGGGCGATTTCGCCAGGCGCATGATGATCTCGGCGCGCATGTGACCCCGCGCG 13024  
552 GCCCGGTTGGGGGCGCTGCGCCCTTACGAGGCGCTTCTGCGCGG 592  
13023 GCCGCGCTGATCTGTGCGCGCTGCGCGAGAACTGCAAGC 12983

RESULT 12  
AF121797 1106 bp DNA linear BCT 21-FEB-2001  
LOCUS Streptomycetes antibioticus A/G-specific adenine glycosylase (muty)  
DEFINITION gene, complete cds.  
ACCESSION AF121797  
VERSION AF121797.1 GI:4455097  
KEYWORDS Streptomycetes antibioticus  
SOURCE Streptomycetes antibioticus  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycetaceae; Streptomycetes.  
REFERENCE 1 (bases 1 to 1106)  
AUTHORS Jones, G.H., Paget, M.S., Chamberlin, L. and Butner, M.J.  
TITLE Sigma-E is required for the production of the antibiotic  
actinomycin in Streptomycetes antibioticus  
JOURNAL Mol. Microbiol. 23 (1), 169-178 (1997)  
MEDLINE 97157951  
PUBMED 9004230  
REFERENCE 2 (bases 1 to 1106)  
AUTHORS Bralley, P. and Jones, G.H.  
TITLE Transcriptional analysis and regulation of the sigma-E gene of

Streptomycetes antibioticus  
Bioclim. Biophys. Acta 1517 (3), 410-415 (2001)  
21240331  
MEDLINE 11342219  
PUBMED 3 (bases 1 to 1106)  
REFERENCE Bralley, P. and Jones, G.H.  
AUTHORS Direct Submission  
TITLE Submitted (19-JUN-1999) Biology, Emory University, 1510 Clifton Rd,  
Atlanta, GA 30322, USA  
JOURNAL

## FEATURES

## source

1..1106

/organism="Streptomycetes antibioticus"

/mol\_type="genomic DNA"

/strain="IMRU 3720"

/db\_xref="taxon:1890"

/clone="pJSE1101"

183..1106

/gene="mutY"

183..1106

/gene="mutY"

/function="DNA mismatch repair"

/note="mutY"

/codon\_start=1

/transl\_table=1

/product="A/G-specific adenine glycosylase"

/protein\_id="AAD1076.1"

/db\_xref="GI:4455098"

/translation="MTAPTKPSSPADVDVSGADLHPVIDMFDHARDPCWRPENG  
AMAVVSEFMIAQTPVSRVLYVYRQWLARVPRDLADAGAEVAMRGYRRL  
RLHGAVALTERHGDVGTGQDLALPGIEYVAAVSPAYQRAHVLDTNRYR  
ARATGVQYPPNATTAERRLARLALDEDEGTARMAAASVELGALVCTANETCVR  
PIAGOCARLAKREHDPGRGQTVAGTGVGKLLAVREHAAPVQBALDQWQ  
EPVDPRALDVLVDPGLVPLRGTYRLPLT"

## ORIGIN

Query Match 13.3%; Score 129.4; DB 1; Length 1106;  
Best Local Similarity 51.8%; Pred. No. 3.4e-08;  
Matches 413; Conservative 0; Mismatches 351; Indels 33; Gaps 4;

20 CCTCTCGCTGTGATCGGGAAGAGCGCGCCCTCTCCCTGGCGG---GGGAGAAAG 76  
253 CGGTATGACATGCTGTTCACAGCAAGCGCGCTGATCGCTCGGGGAGACCCGGAGGG 312  
77 ACCCTTACCGGCTCGTGTCTCGAGGCTCTTCTGACAGACCGCGGTGAGAGAGCGCC 136  
313 GCGGCTGGCGCTTATGCTCAGCGATTCATGCTCAGACAGACCCCGGTAGCGGGTCC 372  
137 TCCCTATTACCGCGCTTCTGAGCGCTTTCACACCTGAAAGCCTTGCGCGGCTT 196  
373 TCCGCTGTACAGAGATGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 432  
197 CCTGGAAGAGTCTTATGAGGCTTGGCAAGGGGGCGGCTTACACCGGGGGGAGAACCC 256  
433 CCCCCGATGAGCGCTCGCGCGGTGGGAGAGCGCTGCGCTACCCCGGGCGCGCTTGGC 492  
257 TCCACGCGCTGCGCGGAGCGTGAAGACTTCCCGGAGCTTCTC-----G 301  
493 TCGACGGCGCGGCTGCGCTTACGAGCGGCGGAGCGGCGGAGCGCTCGGACACAGG 552  
302 CCGAGCTTGGGGGCTTCTGTGCTCGGAGCTTACACCGCGGGCGGCTGCTCATCG 361  
553 CCAGCTCTGGCGCTGCCCGGATCGGAGATACAGACCGCGCGCGCTGCTCTTCT 612  
362 CTTTGGGAGAGGGGTGGCGGCTGAGACGGAAGTCCGAGGGGCTCTCCGCGCTCT 421  
613 CTTAGGGAGGAGGAGCGCGCTTACAGACCAAGTCCGCGGCTTCTGCGCGGCGG 672  
422 TCGC-----CCGGAAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 469  
673 TACAGGGGTGAGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 732  
470 GCTCTCTCCGAGGCGGTGACCGCGGGGTG---TGAACAGAGCGCTCATGAGCTCG 526

Db	733	CCCTGCTCCCGAAGACGAGGGGACCGCCGGCGCTGGGCGCGCGCTCATTGAGCTGG	792
Qy	527	GGGCGACGCTGTGCTCTGCGGAAACGAGCCCGCTTGTGGGGGCTGTGCTCCCTTAGGGCTTTCT	586
Db	793	GCGGCTGTGTGTGACCGCGAAGAACGAGAGTGTGTGCTGTGCTGCTCCCATTCGCCGGGCGAGT	852
Qy	587	GCCGGGGGAAGAGAGGCCCGGGGCGCTACCCGCGCCAGAGAAAGCGCGGGCGAGAGAGG	646
Db	853	GCGCTGGCGGCTGTGGCGGGCAAGCGCGAGACAGACGCGCGCGCGCGGGGCCAGAGCT	912
Qy	647	AGCGCTCTGTGGCCTCTGCTCTCTCTCTGGGCGGAAAGGGGTGTCACTTGAAAAGCTTTAGG	706
Db	913	ACGCGGTACCGACCGGCGAGGTCGGCGCAAGCTGTGCGCTGCTTGAGGAGAGCGCACG	972
Qy	707	GGCGCTTTCAGAGGCTCTACGAGCGTCCCGCTCTTCCCTGAGAGAGCTTCCCGGCGGG	766
Db	973	CGCCCGTCCGACGGAAGCCCTTGACCAAGTGTGTGGCAGAGCGGGTGCAGCGCGCGGG	1032
Qy	767	AGGCGGCTTTCGGGCTG	783
Db	1033	CCCTGACGAGCTGGTG	1049

## RESULT 13

LOCUS	11283 bp	DNA	linear	BCT 12-JUN-2002
DEFINITION	Caulobacter crescentus CB15 section 36 of 359 of the complete			

ACCESSION	AE005710	AE005673
VERSION	AE005710.1	GI:13421519

SOURCE	Caulobacter crescentus CB15
ORGANISM	Caulobacter crescentus CB15

REFERENCE  
1 (bases 1 to 11283)  
Caulobacteraceae; Caulobacter.

Eisen, J., Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R., Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D.

Haft, D.H., Kolonay, J.F., Smilt, J., Craven, M.B., Khouri, H., Shetty, J., Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J., ...

**TITLE**  
Complete genome sequence of *Caulobacter crescentus*  
Praser, C.M.

PUBMED	11259647
MEDLINE	21173698
COSMOS	10000000
LOCUS	10000000

2. (Wassers 1 CO 11203)

**AUTHORS**

Nierman, W. C., Feldblyum, T. V., Paulsen, I. T., Nelson, K. E., Eisen, J., Heidelberg, J. F., Alley, M. R. K., Ohta, N., Maddock, J. R., Potocka, I.

Neelson, M.C., Newcomb, A., Stephens, C., Finkbeiner, N.D., Ely, B., Laub, M.T., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smith, J., Craven, M., Khouri, H., Shetty, J.

Beatty, A., Uccello, J., Iann, A., Moll, A., Valmicheva, U., Ermolaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and Fraser, C.M.

**Direct Submission**  
Submitted (31-JAN-2001) The Institute for Genomic Research, 9712  
Medical Center Dr. Rockville, MD 20850, USA

```
FEATURES
SOURCE
LOCATION/Qualifiers
1. 11283
/organism="Caenorhabditis elegans"

```

```
/moi_type="genomic DNA"  
/strain="CB15"  
/db_xref="taxon:190650"
```

gene	complement(254.763)
CC0365"	/gene="CC0365"
complement(254.763)	

```
/gene="CCU365"  
/note="identified by match to PPM protein family HMM  
PF00410"
```

```
/codon_start=1
```

```
/transl_table=11  
/product -WATN cutthage EQ p subunit"
```

```

/protein_id="AAK22352.1"
/db_xref="GI:13421520"

```

AAKIKAEIDEAQQLREEAOALLADVKQOREDAEROAAMLEAKAKADAKRLAEAEKEL  
EEOIKRRAEMAERKIAQAEAAADVKAAAVDLAAQAAETVLARLAGAKGDTLVDA  
EEOIKRRAEMAERKIAQAEAAADVKAAAVDLAAQAAETVLARLAGAKGDTLVDA

Gene	complement (774. /gene="CC0366"	.1307)
complement (774. /gene="CC0366"	.1307)	

```
/gene="CC0366"  
/note="identified by match to PFAM protein family HMM  
PF0430"
```

```

/codon_start=1
/transl_table=11

```

```
/protein_id="AAK22353.1"
/db_xref="GI:13421521"
```

YAVLSKALLPRVSGAIDERGAKIAGDIADARRMKDEAEQAARAAAAEVAEARAQAQKT  
AADAKARASAEAEQAKEAVLAEKLAABEASIQTARDQAMSQVRVAEETAGAVEAE

```
complement (1434. .1658)
/gene="CC0367"
```

```
/gene="CC0367"  
/note="identified by match to protein family HMM"
```

```
/transl_table=11
/product="ATP synthase F0, C subunit"
```

/db\_xref="GI:13421522"  
/translation="MDAAAAKYIGAGLAMIGAGVGLVFMGNYFQGALRNPTAAA

```
complement(1708, .2469)
/gene="CC0368"
```

```
/gene="CC0368"  
/note="identified by match to protein family HMM"
```

```

/transl_table=11
/product="ATP synthase F0, A subunit"

```

```

/ab_xref="GI:13421523"
/translation="MADPMHQFOIQIVELPTVTVPGLGALDLSITNSVAAMMSALL

```

VOOGMFLVFATSO LAVTLLTGLMVI LTVAVAGFAKNGLKFKL PAPS GVPWPLYLL  
VPIELSP LIRPITLGLR LFGNMLGHHVLKI FAGFVVALGG LGVWGAGALALT SV

```
complement (2475 . 2855)
/ gene="CC0369"
```

```

complement(2775, 12055)
/ gene="CC0369"
/ note="identified by Glimmer2: putative"

```

```

/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

```

/translation="MPKADEPNDKAIESLDARLAAFEAKKAEEKPSKATSEKATODGY
/db_xref="GI:13421524"
/protein_id="P13421.5"

```

KLKLAUIGV L VGVG CMUDDI I A I S E M H I G S S I O D E N I C I V A S H E E D I N G  
APSPSGNADEDTGCGSTAPKQKRD"  
2854: 322

```

/gene=CC0370"
2854 . 3222
/gene=CC0370"

```

```

/notes::identified by Gilmanetz; putative
/codon_start=1
/trans1_table=11

```

```

/product=hypothetical protein
/protein_id="AAK22357.1"
/db_xref="GI:13421525"

```

[illegible]

Query Match	13.1%	Score 127.8	DB 1	Length 11283	
Best Local Similarity	53.6%	Pred. No. 2.7e-06			
Matches 331	Conservative	0	Mismatches 277	Indels 27	Gaps 3
QY	55	CTCCCTGCGCGGGGAGAGAACCTTATCCGCGCTCTGCTCCGAGTCTCTTCCAG	114		
Db	10234	CCCGCGGGGGGGTCCGAGGAGATCCATCCGCGCTCTGCTCCGAGTGAATGCTCCAG	10293		
QY	115	CAGACCCGGGTGAGCAGAGCCCTCCCTATTATCCGCGCTTTTGAGCGCTTTCCACC	174		
Db	10294	CAGACCAAGGTGCGCCACGACCGCCCTATTCTTGAGCTTACCCAGCCCTGCGAGCG	10353		
QY	175	CTGAAGCGCCCTGCGCGCGGCTTCTCCGTGAAGAGTCTTATGGGCTTGACAGGGGGGGG	234		
Db	10354	GCTTATGACCTTCGCGCGGTGAGAGAGCGGACTATATGCGCGCTTGAGGGGCTTTGGC	10411		
QY	235	TACTACCGGCGGGGAGAACCTTCACCGCTGCGCCGAAAGCTGAGAGACTTCCCGG	294		
Db	10414	TACTACGCCCGCGCGCCGACCTTGTGGCTGTGCGCGCGGCTGACCAACATACAGCG	10472		
QY	295	AGCTTC-----GCCGAGCTTCGGGGGCTTCGTGTCTCGGCGCTTACAC	339		
Db	10474	GCGCTTTCCGGGACCGAAGAGGCTTCGCGGCTTCGCGGGCGTGGGGGCTTACACC	10533		
QY	340	GCGGCGGCGGTGAGCTTCATCGCCCTTTCGGGAGCGGAGTGGCGCGGTGAGCGGAACTGC	399		
Db	10534	GCCGCGCGCGGCGCGGATGCGCTTTCGACCGCGCGCAATGTGTGACCGCAATGTC	10593		
QY	400	CGGAGGGCTCTCTCCGCGCTTTCGCGCGGGAGAACCC-----CAAGAGAAAGAG	450		
Db	10594	GAGCGGGGATGTGCGCGCTGTGTGCTGTGAGAGCGCCGATGCCACAGAACCCCGAG	10653		
QY	451	CTTTTCGCGCTTCGCCGAGGCGCTCTCCCGAAGGCGGTGAGACCGCGGGGTGTGAAACGAG	510		
Db	10654	TTGAAGGCGCTTCGCGCGCGGACCTGTGTACCGA---CGATCGCCCGCGGCACTGGGCCGAG	10710		
QY	511	GCCCTCATGAGCTCGGGGCGACGCTGTGCTTCGCGAAACGCGCCCGTTGCGGGGCTTGC	570		
Db	10711	GCGCTGATGATTTTGGGCGCGAACCGTCTGCAAGCCCAAGGCTTCGCTGTGCAACCGCTTGC	10770		
QY	571	CCCTTAAGGGGCTTCTTCGCGGGGAAAGAGAGCCCGGGCGGTGATCCCGCGCGCGAAGAG	630		
Db	10771	CCGGTCTTCGCTCTGTGTGCGGGCTTATGTGTGGCGGGGCGCCCGAGACCTATTCGAGAAAG	10830		
QY	631	CGCCGGCGAAGAGAGCGCCTCTGTGCGCTTCTGTCTTCCTTGCGGCGAGAGGGG	685		
Db	10831	ACCAAGAAAGCGCAGCGCGCGCGCAAGGACGCGGTGAGCTATGTGTGACCGCGG	10885		
RESULT 14					
BX572607/c					
LOCUS	BX572607	345012 bp	DNA	linear	BCT 18-DEC-2003
DEFINITION	Rhodopseudomonas palustris CGA009 complete genome, segment 15/16.				
ACCESSION	BX572607 BX571863				
VERSION	BX572607.1	GI:39651254			
KEYWORDS	complete genomes.				
SOURCE	Rhodopseudomonas palustris CGA009				
ORGANISM	Rhodopseudomonas palustris CGA009				
REFERENCE	Bacteria: Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiales; Rhodopseudomonas.				
AUTHORS	1 (bases 1 to 345012)				
	Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, T.J., Lang, A.S., Tablica, F.R., Gibson, J.L., Hanson, T.E., Torres y Torres, J., Perez, C., Harrison, F.H., Gibson, J. and Haywood, C.S.				
	Complete genome sequence of the metabolically versatile				
TITLE					

JOURNAL  
REFERENCE  
AUTHORS  
CONSRM  
TITLE  
JOURNAL

photosynthetic bacterium *Rhodospseudomonas palustris*  
Unpublished  
2 (bases 1 to 345012)  
Larimer, F.W. and Harwood, C.S.  
*Rhodospseudomonas* genome consortium  
Direct Submission  
Submitted (24-JUL-2003) Submitted on behalf of the *Rhodospseudomonas*  
genome consortium, the DOE Joint Genome Institute, Production  
Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,  
USA, and the Genome Analysis Group, Oak Ridge National Laboratory,  
1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;  
larimerf@ornl.gov

FEATURES  
source

1. 345012  
/organism="Rhodospseudomonas palustris CGA009"  
/mol\_type="genomic DNA"  
/strain="CGA009"  
/db\_xref="taxon:258594"  
280..1221  
/locus\_tag="RPA4336"  
280..1221  
/locus\_tag="RPA4336"  
/function="InterPro IPR001395 COGS COG0667"  
/note="observed by proteomics  
Citation: Proteomics from Verberkmoes et al. (2003)  
unpublished"

gene

CDS

gene  
CDS

1671..2414  
/locus\_tag="RPA4337"  
1671..2414  
/locus\_tag="RPA4337"  
/function="InterPro IPR005834:IPR006402 COGS COG0637"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="putative oxidoreductase"  
/protein\_id="CAE29777.1"  
/db\_xref="GI:39651255"  
/translation="MDIRNLGSGLRVSVGLGNNFGQRTDPTSKVTHKADLGV  
TLPEPTAIYAGMGSESTVGLGRKDIYLAATKPKMATDGTGQASRRYVEAV  
EASLKRKTMDIDYQOHDPLPDIETALADLI,IRAGVRYIGNSPFAPIATA  
EYVARALGTRHYVSGDESVIVARDIEKDLIPAKOYKLGILPFPPLASGLTGKOR  
GTEAPATRRAKMADIRDRYTTDANLIDVTLKPAARGLSLELAFSMACRPQS  
SVIAGATPEQIEONKAVSWKLTAEEMAEVDALITKG"

gene

CDS

gene  
CDS

complement (2341..3021)  
/locus\_tag="RPA4338"  
complement (2341..3021)  
/locus\_tag="RPA4338"  
/function="InterPro IPR001647 COGS COG1309"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="transcriptional regulator, TetR family"  
/protein\_id="CAE29779.1"  
/db\_xref="GI:39651257"  
/translation="MAVRVSVAPKAPPPSPTRDRIYAAALFYSEGRVAVNAE  
KAGVTKTLTYHFRSKDVLVAYLEGKQPNLAFROMFETPGLPKVIGTRNLA  
RAARHPKMGCGFLRTSAELANLPQHPIKIGAAKKFERWLEETPEAGITAEDEL  
ARQWLLIDGSAFVAVLLHRDPSYMETGAAMAAKLVAALPRSFMAARPRAPRTGSG  
SRANSGR"

gene

CDS

gene  
CDS

3103..3894  
/locus\_tag="RPA4339"  
3103..3894

gene

CDS

gene  
CDS

complement (4874..6859)  
/locus\_tag="RPA4341"  
complement (4874..6859)  
/locus\_tag="RPA4341"  
/function="InterPro IPR001279 COGS COG2015"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="Beta-Lactamase-like"  
/protein\_id="CAE29782.1"  
/db\_xref="GI:39651260"  
/translation="MARITPSMTLATTAALSLAAPLAQPNDAEPATRSANVYNS  
LPLADRADFEDAQRGLIASLPDGVPPGAPAPAMDLKPYDFLGDPSPATVNSLR  
QAQNLNLSGLFOVAERYQVRGLDIANVTIVBGTGLI,IDDTLTVTAKAADLYQ  
HRPKPVVALIYHSHIDHFGARGLIDEADAAGKVAAPGLFLEHAAENVIAGN  
AMRRAPQFORGTOLPVGERGVDAAGLAKGVS,IAVNDLIKQYETRSIDGVEI  
EPHIVPSEAPSENIISYPOFKVLNMAEDTHTLHNTYLRGAARICGRLSKYIGBA  
IERGDTVDVIAQHNMPVGRDRVGVYLRKQSDVYKFIDQSRLLNHGLPTETIAB  
RLTLPELSTEEFARGGYGSVSHNAKVQFYLGWYDANPADLPLPRAQAKKEIDY  
MGGAAYLIAARDYKAGQRYAVATVSKLVFADPATTEALAGADLEQGXOAEIS  
TMRNAVYLIAGQELRNLKIDTSVTSNDLKGVSDIADFPFLAVLNAAKEGGIIV  
NMTPTDKERTYMWLNSAL,THISGKSDNADSVTLNFTFDALISLKQGFCAVVS  
GDLMVSGNPLKRLBELFGLFEDFSBNFEVIEPVAKAKE"

gene

CDS

sig\_peptide

gene  
CDS

7091..8143  
/locus\_tag="RPA4342"  
7091..8143  
/locus\_tag="RPA4342"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="CAE29783.1"  
/db\_xref="GI:39651261"  
/translation="MRYSAEDLQAAAGVITPAQEAALPAFLDARRPAAASDAFED  
IVHLWYAGALIVISAMGLFTMAFAGLGGAALTATLTIYAVAFWMTGHLVHYNLR

gene

CDS

gene  
CDS

7091..8143  
/locus\_tag="RPA4342"  
7091..8143  
/locus\_tag="RPA4342"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="CAE29783.1"  
/db\_xref="GI:39651261"  
/translation="MRYSAEDLQAAAGVITPAQEAALPAFLDARRPAAASDAFED  
IVHLWYAGALIVISAMGLFTMAFAGLGGAALTATLTIYAVAFWMTGHLVHYNLR

TPGGLCIAVAVGMPLAVYVOBAGFAGLSEFGKPDTRSFYIMVKGSFIFMETATIVA  
ALVALAFYRPFPIVLMVALMPLISMDLIPMTITGPHGDFKRVKSGFGLVMTIAA  
VIVHLRORSGDFAFMLYLFVGVNTFMGRTATSNGRITDAULICANWLVRLGIAVYIGR  
RFAVAFGALGIAIYIGDLAEKLFROSLFEPALSLGIALIGLFLHRRLPALIAEAWC  
DRLPKGMORLEVPAA"  
order(7232, .7300, 7313, .7381, 7400, .7459, 7517, .7585,  
7604, .7663, 7706, .7759, 7778, .7846, 7889, .7957, 7994, .8062)  
/locus\_tag="RPA4342"  
/note="9 probable transmembrane helices predicted  
byTMHMM2.0"  
8650, .11238  
/gene="uvrB"  
/locus\_tag="RPA4343"  
8650, .11238  
/gene="uvrB"  
/locus\_tag="RPA4343"  
/function="interPro  
IPR001410:IPR001650:IPR001687:IPR001943:IPR004807 COGS  
COG0556"  
/note="observed by proteomics  
Citation: Proteomics from Verberkmoes et al. (2003)  
unpublished"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="possible excinuclease ABC subunit B"  
/protein\_id="CAE29784.1"  
/db\_xref="GI:39651262"  
/translation="MEALGVKATAEALLESILRDGRPEFGEDGVYLVWVPRPREK  
SEGVRFLKSDYOPRGDQPTAKELVSGIDENDRTOVLGVTSKGTVMKIVLEAT  
ORPATILANPKTLAOLYGEFNPFPDPAWVEVSGYDTEAVYPRPDYITEDDS  
INQIDRMHNAATRALERDDYIVASVCTYIGISGVETTYTANPAKXGERIDOROL  
INDLVALQYKRTQADSRGTFRVRGDVIDIFPAHVEDRMRKVMDEBIEGIEEDPL  
TGHKDELEFVKIYANSHVYTPRLIQAIQSIKTELKRLQDLAOGSLAQLBQ  
RTTFPIEMMEATGSCAGIENYSRYLGRGPPPTLFEYVDNALVFADBSHVSIPQ  
IGAMPKDFRRKATLAEGFRLPSCMDNPLFEEDMMRPPTVAVASAPPAEINES  
GGFVPEOVRAPGLIDPVDIRPARTOVLDIGEVATATAGVARTILTYLTRMAEDL  
TEFLHQCGRVRYMHSIDITIRIETIRRLRGAEDALVGINLBSGLDIPCALVAT  
LDADKEFLRSETSLQITIGRAARNVDGVIIYADQMTGSMORSIDETIRREKQIET

Query Match 12.8%; Score 124.4; DB 1; Length 345012;  
Best Local Similarity 51.9%; Pred. No. 2,7e-08;  
Matches 377; Conservative 0; Mismatches 316; Indels 33; Gaps 3;

2 TGGAGGCTTGGCGGAAAGCCCTCTGCTGCTGATCCGGGAAACGCCGCCCTCCCT 61  
173098 TCGCGCGCGCGCGCGCGCTGCTGCTGCTGATCCGGGATCCGCCCTGCGCT 173039  
62 GCGCGCGCGGAG-----AGGAGCCCTTACCGCGCTGCTGCTGCGAGTCCCTTC 109  
173038 GCGCGCGCGCGCGCGCGCGATCCCTATGCGGTGTGCTGTGCGAGATCATGC 172979  
110 TGCAGAGAGCCCGGCTGAGACAGCCCTCCCTATTACCGCGCTTCTTGAGCGCTTTC 169  
172978 TGCAGAGAGACACCGTTCGTGGGGGTGGGCCGATTTGACAAAGTTCAATGGCGGTGCG 172919  
170 CCAACCTTGAAGCCCTGCGCGCGCTTCCCTGGAAGAGTCTTGAAGTCTGGACGGGG 229  
172918 CCAACCGTGAAGCGCGCTGCGCGCGCTCCCTGCAAGACCTGCTGAAGATGTGGCGCGGC 172859  
230 CCGGCTACTACCGCGCGCGCGGAAACACTCCACCGCGCTGCGCGGAGCGTGAAGAGC--- 286  
172858 TCGGCTACTACTACCGCGCGCGCAATCTCAAGCGCTGCGCGGTGGCGGTGAAGCGCCACG 172799  
287 -----TTCCCGGAGCTTTCGCGAGCTTTCGCGGAGCTTCTCTGCTCTCGAGCTT 334  
172798 ATGAGAGCGCGCTTTCGCGAGACCGAGAGAGGGGTGCGGCGCGCTCGCGCGCTGCGGCGCT 172739  
335 ACAACCGCGCGCGCGGTGCGCTTCATGCTTTCGCGGAGCGCGGTGCGCGCGGTGAAGCGGA 394  
172738 ACAACGCGCGCGCGCGATGCGCGGATTCGCTTCAAGCGCGCGGAGCGATGCGCGGTGAAGCGGA 172679  
395 AGCTCGGAGAGGATCTCTCCGCTTTCGCGCGCGGAGAAAGC-----CCCAAGAGAGAGG 448

Db 172678 ATATGAGACGGGGTGTGTGTGCGGGCTGTACGGCGTTCAGAGACGAACCTGCCGAAGCGCAAGC 172619  
Qy 449 AGCTTTTGGCCCTCGCCAGGAGCCCTCTCCCGAGAGGCGTGTAGACCCGGGGGTGTGAACC 508  
Db 172618 CGCGCATCAAGCGCGCTGCGCGAGACCCCTGCTCGGCGCCCTCGCGGTACAGCGGCC 172559  
Qy 509 AGCGCCATGAGAGTGTGGGGGCAACGCTGTGCTGCGCGAAAGCGGCCCTTGTGGGGGCT 568  
Db 172558 AGCGCCATGAGATTTGCGCGGCACATCTTCAGCGCGCAAGAAAGCGGCTTGTGGCGCTGT 172499  
Qy 569 GCGCCCTGAGGGGCTTCTGCTGCGGGGAGAGAGCGCCCGCGCGCTACCCCGCCAGGA 628  
Db 172498 GCGCGCTATGACAGGCGCTGCACCGACGCGCTGCGGGAGATCCGAGAGACTTTCCGCGCA 172439  
Qy 629 AGCGCGGCGCAAGAGAGAGCGCTGCTGCGCGCTTCTCTGCGCGGAGAGGGGCTGC 688  
Db 172438 AGCGCGCGAGAGAGACCGGAGACGCTGCGCGCGCGCTTGTGTGATCCGCGCGC 172379  
Qy 689 ACCTGG 694  
Db 172378 ATCAAG 172373

RESULT 15  
AE005065/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
2 (bases 1 to 11034)  
Ng,M.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,  
Shukla,H.D., Laeky,S.R., Balliga,N., Thorsen,V., Sbrogna,J.,  
Swartzell,S., Weir,D., Hall,U., Dahl,T.A., Weir,D., Goo,Y.A.,  
Leithausen,B., Keller,K., Cruz,R., Danon,M.J., Hough,D.W.,  
Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angvine,C.M., Dale,H.,  
Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angvine,C.M., Dale,H.,  
Isenbarger,T.A., Peck,R.F., Pohlischod,M., Spudich,J.L.,  
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,  
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and  
Dasarma,S.  
Genome sequence of Halobacterium species NRC-1  
Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)  
20504483  
11016950  
2 (bases 1 to 11034)  
Ng,M.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,  
Shukla,H.D., Laeky,S.R., Balliga,N., Thorsen,V., Sbrogna,J.,  
Swartzell,S., Weir,D., Hall,U., Dahl,T.A., Weir,D., Goo,Y.A.,  
Leithausen,B., Keller,K., Cruz,R., Danon,M.J., Hough,D.W.,  
Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angvine,C.M., Dale,H.,  
Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angvine,C.M., Dale,H.,  
Isenbarger,T.A., Peck,R.F., Pohlischod,M., Spudich,J.L.,  
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,  
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and  
Dasarma,S.  
Direct Submission  
Submitted (14-JUN-2000) Institute for Systems Biology, 4225  
Roosevelt Way NE, Seattle, WA 98105, USA  
Location/Qualifiers  
1, .11034  
/organism="Halobacterium sp. NRC-1"  
/mol\_type="Genomic DNA"  
/strain="NRC-1"  
/db\_xref="taxon:64091"  
complement(70, .1719)  
/gene="VNG510C"  
complement(70, .1719)  
/gene="VNG510C"  
/note="conserved hypothetical protein"

```
/codon_start=1
/transl_table=11
/db_xref="Vng1510c"
/protein_id="AAG19804.1"
/db_xref="GI:10581002"
/translation="MSSNADDAQAQAACEELVSRILAGVCKDDIEDAKLDGCEYAA
SKVPHSDLDLVADGRDELVDLORQKPRVATGAVPVALMTSPHCTPHKCLYCPG
GPAEFTSSOSYTGHEPAAAGVONDDPVGQVTLRLHOLQIGHPVKXELLIMCGT
MTASHDYQEFVYKALQALNDPDPDEPAPNQTSPADPAEDFRLIEVIANDEH
GALNNIGTFETKPDWCDEQVDRMLDGAATKVGVQVTFERVNRKMGHGVQASV
DANRLRDAAFKQFHMMPGQPMGSKEMVLEDFRRLPDDERMKPDYLKIPTLVVRDT
ITYDMNRGDEFEPLHNEADLVAEIKMIPRYRLQVRQIDIPADFIADGVMKSNLR
OLARORAEHGYTCDCIRCREAGNDDPENASLDVETAAAGGGEOPISIEDQOOL
LVGFARLRFPRPEPARBELTDALRLBELHVGSEVSGEDDGDHOGYGRQLMERA
EHLAADSGFEKLSYISGIGARQYKQQLGHQDGPYYSKRL"
complement(1767..3821)
/gene="VNG1511C"
complement(1767..3821)
/cds
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1511c"
/db_xref="GI:10581003"
/translation="MQMATCIYCGSSSTDGHCATHEEDVVPDFRGDSFDQLTGGRFPR
GTVDGFABFVGVVGGTVTGLHRSVEVPGVDSLAMVGDDVFVQVVDVNDQNLIDL
GMSIRBSMRDFRGLVHDPANDSDAALPADDEADSDTDDGTTDSDSDAETAGGRK
VSDADATTERAADASFESGADDDDTHEITVRVDLDDHVDADVTEGEITRAAQTSQ
PYFELTDETCADVAAAFVAGVAPPEIEGAVVVLGTVERRRGLQVETEAUTEL
TGDRADVRNMDAALDDBRAPATALLAPDDADVAHDDIYDAATPRAPAVIDAR
VTVNPTTVSGYVAGTAIERALLPLRIDEAREBAEHYVDRPLDADFIYDIAAGTD
VTSMLAAHRHDERKPLFVFGSTSTESTDALDLIDYADTVVVDGVDGADDAAGMD
YLSVPTAAGEPVTGALQAQAAVINDREDIAHLPAVAVWTDLPAAVADLAADS
YSDARALRDVALEAFYOSGEDKRELIADLKGADDDLAHIAIQFRRLATELS
TAEPHLSTRHDDGVAFETLDVYATGHQDPRPVLLDLALHRAADSTVDLIGAGEDAL
RVRSDDAVDVAAVDAIDDDLTTAGTARABGRLEFLAGERDAVIDAVDAIQAQ
LA"
complement(3867..4103)
/gene="VNG1513H"
complement(3867..4103)
/gene="VNG1513H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1513h"
/protein_id="AAG19806.1"
/db_xref="GI:10581004"
/translation="MVELKDSIKVNGPFTNKHADAKLFVDVDATEAREGDKRVLFVE
DDSGNEVQVALFPEQVALLADDIDALTRDSVFPE"
complement(4154..4762)
/gene="VNG1514H"
complement(4154..4762)
/gene="VNG1514H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1514h"
/protein_id="AAG19807.1"
/db_xref="GI:10581005"
/translation="MTTWVEPBGGREREPGLIAGSFYOVLNPNVSPFDEAVSPDDQAP
GLVAVAVVGVAAATELAATPSDALAPPTAOMLAVALVAVLVLPAAVHAALSAIQ
TLCIVATISISRGVSEVQLIYAAAPCVVGTIGVPLVTLTGVMWACLLVVGTSV
VHESILPRAVIAAAPVIAFGVIAAGFGAIPAGEQYTHAAGLAF"
complement(4766..5353)
/gene="tdk"
/note="synonym: VNG1515G"
complement(4766..5353)
/gene="tdk"
/note="tdk"
/codon_start=1
/transl_table=11
/product="thymidine kinase"
```

```
/protein_id="AAG19808.1"
/db_xref="GI:10581006"
/translation="MHAITNSGWVEVLTGMSFKTEELLRLRRAETAGODVAATP
AVDDRYGEBALIGHAGRSMAATVEPTABEYASIPTLINGQVVAIDANFPPELVUD
VCOELADGRVRLVSGTDQTFRGSPFPPQMLAIEYVDKRAICMQCGSPATRNOR
LIEGEPAHYDDPTVMVAEETYEARCRRCHVVRRE"
5460..5900
/gene="VNG1518H"
5460..5900
/gene="VNG1518H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1518h"
/protein_id="AAG19809.1"
/db_xref="GI:10581007"
/translation="MYDDILLPTDGSAAADAANAVIDLASKYNATLHVLYADTTEY
STVFEDAVTDPLEREGERVVDVAABADGAVAMGVILHGVFETIOAYVDEGID
TVWGTGRRGLGRALLGSVTERVVRTSDVPLTVRETDGGA"
complement(5917..6126)
/gene="VNG1519H"
complement(5917..6126)
/gene="VNG1519H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1519h"
/protein_id="AAG19810.1"
/db_xref="GI:10581008"
/translation="MEHNVGMDALRLILGAVLGSAAATLTIAGYGGPSALTPALGV
GALLILASGMNTCGVRYLVGVDTT"
complement(6203..7141)
/gene="muty"
/note="synonym: VNG1520G"
complement(6203..7141)
/gene="muty"
/note="muty"
/codon_start=1
/transl_table=11
/product="A/G specific adenine glycoylase, repair
protein"
/protein_id="AAG19811.1"
/db_xref="GI:10581009"
/translation="MTTGDSGRAGATGPADTTALQTLADVWYTDHSRSPFWRETDP
YEILVSEVMSQQTQLSRVIDAMRAFLRWPTTALAADRSDDVGFWSHSLGYNRA
THLHEAQOVEDTDGAIPTPTAPLSEISMGVPTAAVASFANAGNAVVDTVNRY
LYRAFEGIRDDDDPDYRPLANELLPDGSRVMNNAVNEIGAVACQTPRCDEAECPLR
EMCHAYOTGDPTADVPVTOPSFEGSRQFRGRIVRLIGHEDEMELDALGHRVVRDYTP
DGEYGRMWGLSLDLADDGLVRYEYNGERYVVALE"
7627..9558
/gene="Htr8"
/note="synonym: VNG1523G"
7627..9558
/gene="Htr8"
/note="Htr8"
/codon_start=1
/transl_table=11
/product="Htr8 transducer"
/protein_id="AAG19812.1"
/db_xref="GI:10581010"
/translation="MGRSARGSGAETTGGLAAAVKTYLETPPGESTIPAAWRRHR
NVRVFLVAHLPLILALULYGTESAVTGATIPPTPGILIAELGIVGLVGLASIPSV
SRRRRTLAISRAVLAASSVIVQFSGFIKAHFHFGMAVAVVEDMLPFLALGLVYV
FTTHGFEMIAERYVHTTAIINNPMWCGIHGAFVLILAGLALMNYSTERSRASOK
RLBEARQAOQVEDLEBAROEIEAEKAEARKADAEAREAAAOOREVALNARLE
ATANTGAAARADGDLVRLDDVENDMAAIIAASFNEMLDETETTIREIOVASD
VAAASEPADAGVETIEDASQVSETVQEIYAAGDEQEKLETVSGEMTDSAAIEVA
ASADSVAERSHETNAVAGDSQTRBOAIAUSRTVQASVESTVQVNEALDDQIAISFI
VDLISDAEQTNMLANANI EAARADS GGFVAVADVESTVQVNEALDDQIAISFI
DIDAKQATTEARTADESVODAISAVDAVDAFGTVAENAEETDGOEISTTTDDQ
AASTEBAVSMIAEVSIDISTATAADAAQOASTAEOQTAAATISNTALRQADRLQG
LVSTFDVHDSASTAARSE"
complement(9562..10938)
```

CDS  
/gene="VNG1524C"  
complement(9562, .10938)  
/gene="VNG1524C"  
/note="conserved hypothetical protein"

/codon\_start=1  
/transl\_table=1  
/product="Vng1524C"

Query Match 12.7%; Score 123.8; DB 1; Length 11034;  
Best Local Similarity 53.9%; Pred. No. 9.4e-08;  
Matches 313; Conservative 0; Mismatches 247; Indels 21; Gaps 2;

```
QY 7 GCTTGGCGGAAAGCCCTCTTCCGCTGTATCCGGGAAAAGCCCGCCCTCCCTGGCGG 66
DB 7084 GCCCTCCAGACCGCACTCGTCACTGTATCACCGACTCCACCGGTCTTCCCGTGGCG 7025
QY 67 GGGGAGAAAGACCCCTTACCGCGTCTGTCTCCGAGTCTTCTGAGAGACCCGGGTG 126
DB 7024 GAGACCAACGACCCCTTACGAGATCTGTCTGTGAGGTATGAGCCAGACCAAGCTC 6965
QY 127 GAGCAGGCGCTCCCTATTACCGCGCTTCTGTGAGCGCTTCCACCTGAAGGCTTG 186
DB 6964 TCCCGGGTATCGACGGCGGTGCGGGCTTCTCGATGGTGGCCGACACCGCGGTTG 6905
QY 187 GCGGCGCTTCCCTGGAAGGTCTTAAAGTCTG-----CAGGGGCGGCTTACTAC 240
DB 6904 GCGGCGCGCGACCGACGACGCTGTGGTCTGTGCGCACAGCTCGGTTACAC 6845
QY 241 CGGCGGCGGAAACACTCCACCGCTGCGCCGAAAGCTGA-----GGAG 285
DB 6844 AACCGCGCACTCACTCAAGAGCGGCCCAACAGTGAACGACTACGACGCGCG 6785
QY 286 CTTCCCGCGAGCTTGGCGAGCTTGGGGGCTTCTGTCTGGGCTTACACCGCGCG 345
DB 6784 ATCCGCGGAGCGCGCGAGCTCTGAGCTCATGGCGCTGCGCCGTTACACCGCAAC 6725
QY 346 GGGGTGGCTTCATCGCCTTCCGGGAGCGGTTGCGCGGTGAGCGGAAAGTCCGAGG 405
DB 6724 GCGGTGGCGTCTTCTGCTTCAAGCGCGCAACGCGTGTGACACGACGTCAAGCGC 6665
QY 406 GTCTCTCCGCGCTTCTGCGCCCGGAAAGCCCAAGAGAGAGCTTTTCCGCTCGCC 465
DB 6664 GTGCTTACCGCGCGCTTCAAGGAGATCCGGAAGAGACGACCGGATTAACGGCGCGT 6605
QY 466 CAGGCGCTCTTCCCGAGAGCGGTGAACCGGGGGGTGTGAACGAGCCTCATGAGCTC 525
DB 6604 GCCAAGAGTGTGTGCCGACGAGCATGCCGCGTGTGAACACCGCGTATGAATCTG 6545
QY 526 GGGGCCACGCTGTGCTGCGGAAACGGCCCGCTTGGGGGGC 566
DB 6544 GGGGCGGTGCTGTGCAAGACCCCGCTGTGACGAGGC 6504
```

Search completed: March 4, 2004, 05:06:40  
Job time : 2827 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2004, 03:00:31 ; Search time 94 Seconds  
(without alignments)  
5756.140 Million cell updates/sec

Title: • US-09-938-901A-1

```

Sequence: 1 atggaagcctggcgaagc.....tagtccccctcccgagca 975

```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 segs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

```
Database : Issued Patients NA:*
1: /csgn2_6/p/codata/2/ina/5A COMB. seg.*
2: /csgn2_6/p/codata/2/ina/5B COMB. seg.*
3: /csgn2_6/p/codata/2/ina/6A COMB. seg.*
4: /csgn2_6/p/codata/2/ina/6B COMB. seg.*
5: /csgn2_6/p/codata/2/ina/PCCTS. seg.*
6: /csgn2_6/p/codata/2/ina/Backfillse1. seg.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	125.4	12.9	963	4	US-09-352-991A-6832	Sequence 6832, Ap
2	125.4	12.9	1401	4	US-09-352-991A-6886	Sequence 6886, Ap
3	125.4	12.9	1413	4	US-09-352-991A-6739	Sequence 6739, Ap
4	113	11.6	1551	4	US-09-489-039A-4788	Sequence 4788, Ap
5	107.2	11.0	549	4	US-09-352-991A-6783	Sequence 6783, Ap
6	104	10.7	1858	4	US-08-813-574-1	Sequence 1, Appl
7	87.2	8.9	2299	4	US-09-651-656-38	Sequence 38, Appl
8	87.2	8.9	2293	4	US-09-650-855-38	Sequence 38, Appl
9	86.6	8.9	4403765	3	US-09-103-840A-2	Sequence 2, Appl
10	86.6	8.9	4411529	3	US-09-103-840A-1	Sequence 1, Appl
11	82	8.4	9909	4	US-08-9611-527-12	Sequence 12, Appl
12	63.6	6.5	1212	4	US-09-107-532A-3654	Sequence 3654, Appl
13	62.4	6.4	1188	4	US-09-134-000C-1110	Sequence 1710, Ap
14	62.4	6.4	1227	3	US-09-385-028-23	Sequence 23, Appl
15	62.2	6.4	1227	4	US-09-726-614-23	Sequence 23, Appl
16	62.2	6.4	1227	4	US-09-385-040-23	Sequence 23, Appl
17	62.2	6.4	15079	3	US-09-385-028-1	Sequence 1, Appl
18	62.2	6.4	15079	4	US-09-726-614-1	Sequence 1, Appl
19	62.2	6.4	15120	4	US-09-385-040-1	Sequence 1, Appl
20	61.8	6.3	1044	4	US-09-535-681A-1143	Sequence 1143, Ap
21	61.4	6.3	1230025	4	US-09-198-452A-1	Sequence 1, Appl
22	59.6	6.1	20235	1	US-07-642-734C-3	Sequence 3, Appl
23	58.6	6.1	20235	3	US-08-439-009A-3	Sequence 3, Appl
24	58.8	6.0	1041	4	US-09-328-352-856	Sequence 856, Appl
25	56.6	5.8	2889	1	US-08-537-002A-4	Sequence 4, Appl
26	56.6	5.8	2889	3	US-08-863-010-4	Sequence 4, Appl
27	56.6	5.8	2889	3	US-09-024-429-4	Sequence 4, Appl

## ALIGNMENTS

28	56.6	5.8	3600	1	US-08-537-002A-5	Sequence 5, Appl 1
29	56.6	5.8	3600	3	US-08-863-010-5	Sequence 5, Appl 1
30	56.6	5.8	3600	3	US-09-024-429-5	Sequence 5, Appl 1
31	55.8	5.7	2681	3	US-08-928-213B-7	Sequence 7, Appl 1
32	55.6	5.7	2363	4	US-09-818-780-22	Sequence 22, Appl 1
33	55.6	5.7	5300	4	US-09-266-965-76	Sequence 76, Appl 1
34	55.2	5.7	2634	3	US-08-941-936-1	Sequence 1, Appl 1
35	54.8	5.6	1725	4	US-09-225-991A-1	Sequence 1, Appl 1
36	54.8	5.6	2008	4	US-09-225-991A-1068	Sequence 1068, Appl 1
37	54.8	5.6	2586	4	US-09-225-991A-1804	Sequence 1804, Appl 1
38	54.8	5.6	3378	4	US-09-225-991A-2158	Sequence 2158, Appl 1
39	53.8	5.5	47981	4	US-09-679-479-1	Sequence 1, Appl 1
40	52.8	5.4	4257	2	US-08-657-473-1	Sequence 1, Appl 1
41	52.8	5.4	4257	3	US-09-225-821A-1	Sequence 1, Appl 1
42	52.8	5.4	4257	3	US-08-844-639-1	Sequence 1, Appl 1
43	52.8	5.4	12001	1	US-08-456-568A-11	Sequence 11, Appl 1
44	52.6	5.4	2508	4	US-09-755-282B-48	Sequence 48, Appl 1
45	52.6	5.4	3147	2	US-08-781-802-7	Sequence 7, Appl 1

```

RESULT 1
US-09-252-991A-6832
; Sequence 6832, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 31142
; SEQ ID NO 6832
;
; LENGTH: 963
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6832

```

Query Match	12.9%	Score 125.4	DB 4	Length 963
Best Local Similarity	54.7%	Pred. No. 2.1e-16		
Matches 311	Conservative	0	Mismatches 231	Indels 27
			Gaps	2
QY	CTCCCTCGCGGGGGGAGAAAGACCCCTTACCGCGGCTGGTCTCCGAGTCTCTTGAG	114		
DB	CTGCCCTTGCGACAGGAGCATACCCCTTACCGGGGTGGGTCTCGGAATATGCTCGAG	112		
QY	CAGACCCGGGTGAGACAGCCCTCCCTCTATTACCGCGCTTTCGAGCGCTTCCACC	174		
DB	CAGACCCAGGTTCAGACACGCTGCTGGTACTTTCAGACGTTTCATGAGCAGCTTGCAGC	172		
QY	CTGAAGGCGCTTGCGCGCGGCTTCCCTGGAAGAGTCTTAAAGGTCTGCGAGGGGGCGGG	234		
DB	GTCGAGGCACTGCGCGCGCGCGCCAGGACGMAAGTCTGCACTGTGTGACCGGGCTCGGC	232		
QY	TACTTACCGCGGGGCGGACACCTCTCCACGCGCTGGCGCGAAGGCT-----G	279		
DB	TACTACACCCGTCGCGCGCACTCTGACAAAGCAGCGCGCAGATGTGTGTGAGCGGCATGCG	292		
QY	GAGAGACTTCCCGCCAGCTTGCGCCAGCTTGGGGGCTTCTGTGCTCGGCGCTTACAC	339		
DB	GAGAGACTTCCCGCGCGAGCTTGCGCAACTGTGACAAAGCAGCGCGCAGATGTGTGTGAGCGGCATGCG	292		
QY	GCGGAGTTCCTCCCGCGAGCTGAGGAACTCGCGGAACTGCGCGGATCGGCGCTTCAAC	352		
DB	GCGGAGTTCCTCCCGCGAGCTGAGGAACTCGCGGAACTGCGCGGATCGGCGCTTCAAC	352		
QY	GCGGCGGCGGTGCGCTTCATGCTCTTTCGCGGAGCGGGTGGCGCGCGGTGAGCGGGAACGTG	399		
DB	GCTGAGAGCATGCGCAGCTGTGTGATGTGGGCTTGCAGCGACCATCTCTGACGCGCAACGTG	412		

QY 400 CGAGGAGTCTCTCCGCTCTTGCCTCCGGAAGCCCAAGAGAGACTTTTCGCC 459  
DB 413 AAGCGGTACTGCGCGCTACTGCGCAGAGACGGCTATCCCGCAACCGAAGTGGCC 472  
QY 460 CTGCCCCAGGCGCTCTCTCCCGAGGCGGTGAGACCGCGGGTNG-----AAC 507  
DB 473 AGGCGCTGTGGAGAGCCCGGAACGCTTACCCCGACACGAGGTCAACACTACAC 532  
QY 508 CAGGCTCATGAGAGCTCGGGGACAGGTCTGCTCGGAAAGCGCCCGTTGCGGGGCC 567  
DB 533 CAGGCGATATGAGACTCTCGGCGCAACCTCTGCAAGCGAGCAAGCCAGTTGCTGCTT 592  
QY 568 TGCCCCCTAGGGGCTTTCGCCGGGGA 596  
DB 593 TGCCCGTGTCTCCGCTGCGCGCGCA 621

## RESULT 2

US-09-252-991A-6886/c  
Sequence 6886, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6886  
LENGTH: 1401  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6886

Query Match 12.9%; Score 125.4; DB 4; Length 1401;  
Best Local Similarity 54.7%; Pred. No. 2.2e-16;  
Matches 311; Conservative 0; Mismatches 231; Indels 27; Gaps 2;

QY 55 CTCCTCTGCGGGGAGAGAGACCTTACCGCGCTCTGCTCGAGGTCTTCTGACG 114  
DB 1137 CTGCTCTGCAAGAGGATACACCTTACCGGGTGTGGTCTGGAATATATCTGACG 1078  
QY 115 CAGACCGGCTGAGAGAGGCTCTCCCTATTACCGCGCTTCTGAGCGCTTCCGACC 174  
DB 1077 CAGACCGGCTGAGAGGCTCTCCCTATTACCGCGCTTCTGAGCGCTTCCGACC 1018  
QY 175 CTGAGGCTCTGCGCGGCTTCCCTGAGAGAGTCTTGAAGTCTGCGAGGGGCGGGC 234  
DB 1017 GTGAGGCTCTGCGCGGCTTCCCTGAGAGAGTCTTGAAGTCTGCGAGGGGCGGGC 958  
QY 235 TACTACCGGCGGAGGAGACCTTCAACCGCTGCGCGAGCGT-----G 279  
DB 957 TACTACCGGCTGCGCGGAGGAGACCTTCAACCGGCGAGTCTGTGAGAGCGGATCG 898  
QY 280 GAGAGGCTCTTCCCGAGGCTTCCCGAGGCTTCCCGAGGCTTCTGCTGCGGCTTCAAC 339  
DB 897 GGGAGGCTTCCCGCGGAGGCTTCCCGAGGCTTCCCGAGGCTTCCCGAGGCTTCCCGAG 838  
QY 340 GCGGCGGCTGAGGCTTCAACGCTTCCCGAGGAGCGGCTGCGGCGGCTGAGAGGAGTGC 399  
DB 837 GCTGAGGCTTCCCGCGGAGGCTTCAACGCTTCCCGAGGAGCGGCTGAGAGGAGTGC 778  
QY 400 CGAGGAGTCTCTCCCGCTTCTGCGCGGAGAGGCTTCCCGAGGAGGAGGAGGAGGAGGAGG 459  
DB 777 AAGCGCTGAGGAGGCTTCTGCGCGGAGAGGCTTCCCGAGGAGGAGGAGGAGGAGGAGG 718  
QY 460 CTGCGGAGGAGGCTTCTCCCGAGGAGGCTTCCCGAGGAGGAGGAGGAGGAGGAGGAGG 507

DB 717 AGGCGCTGTGGAGAGCGCGGAGAGGCTTCAACCGCGAGCAGCGGTCAACACTACACC 658  
QY 508 CAGGCTCATGAGAGCTGCGGGGCAACGCTCTGCTCTGCGGAAAGCGCCCGTTGCGGGGCC 567  
DB 657 CAGGCGATATGAGAGCTTCTGCGCGCAACCTCTGCAAGCGGAGCAAGCCAGTTGCTGCTT 598  
QY 568 TGCCCCCTAGGGGCTTTCGCCGGGGA 596  
DB 597 TGCCCGTGTCTCCGCTGCGCGCGCA 569

## RESULT 3

US-09-252-991A-6739  
Sequence 6739, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6739  
LENGTH: 1413  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6739

Query Match 12.9%; Score 125.4; DB 4; Length 1413;  
Best Local Similarity 54.7%; Pred. No. 2.2e-16;  
Matches 311; Conservative 0; Mismatches 231; Indels 27; Gaps 2;

QY 55 CTCCTCTGCGGGGAGAGAGACCTTACCGCGCTCTGCTCTGAGGTCTTCTGACG 114  
DB 409 CTGCTCTGCAAGAGGATACACCTTACCGGGTGTGGTCTCGGAATATATCTGACG 468  
QY 115 CAGACCGGCTGAGAGGCTCTCCCTATTACCGCGCTTCTGAGCGCTTCCGACC 174  
DB 469 CAGACCGGCTGAGAGGCTCTCCCTATTACCGCGCTTCTGAGCGCTTCCGACC 528  
QY 175 CTGAGGCTCTGCGCGGCTTCCCTGAGAGAGTCTTGAAGTCTGCGAGGGGCGGGC 234  
DB 529 GTGAGGCTCTGCGCGGCTTCCCTGAGAGAGTCTTGAAGTCTGCGAGGGGCGGGC 588  
QY 235 TACTACCGGCGGAGGAGACCTTCAACCGCTGCGCGAGCGT-----G 279  
DB 589 TACTACCGGCTGCGCGGAGGAGACCTTCAACCGGCGAGTCTGTGAGCGGATCG 648  
QY 280 GAGAGCTTCCCGAGGCTTCCCGAGGCTTCCCGAGGCTTCTGCTGCGGCTTCAAC 339  
DB 649 GGGAGTCTCCCGCGAGGCTTCCCGAGGCTTCCCGAGGCTTCCCGAGGCTTCCCGAG 708  
QY 340 GCGGCGGCTGAGGCTTCAACGCTTCCCGAGGAGCGGCTGCGGCGGCTGAGAGGAGTGC 399  
DB 709 GCTGAGGCTTCCCGCGGAGGCTTCAACGCTTCCCGAGGAGCGGCTGAGAGGAGTGC 768  
QY 400 CGAGGAGTCTCTCCCGCTTCTGCGCGGAGAGGCTTCCCGAGGAGGAGGAGGAGGAGGAGG 459  
DB 769 AAGCGCTGAGGAGGCTTCTGCGCGGAGAGGCTTCCCGAGGAGGAGGAGGAGGAGGAGG 828  
QY 460 CTGCGGAGGAGGCTTCTCCCGAGGAGGCTTCCCGAGGAGGAGGAGGAGGAGGAGGAGG 507  
DB 829 AGGCGCTGTGGAGAGCGCGGAGGCTTCAACCGGAGGAGGAGGAGGAGGAGGAGGAGG 888  
QY 508 CAGGCTCATGAGAGCTGCGGGGCAACGCTCTGCTCTGCGGAAAGCGCCCGTTGCGGGGCC 567  
DB 889 CAGGCGATATGAGAGCTTCCCGCGGAGGCTTCTGAGAGGAGGAGGAGGAGGAGGAGGAGG 948

QY 568 TGCCCCCTAGGGGCTTTGCGCGGGGAA 596  
Db 949 TGCCCCGTGTCTCCGCTGCGCGCGCA 977

RESULT 4  
US-09-489-039A-4788  
Sequence 4788, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489, 039A  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 4788  
LENGTH: 1551  
TYPE: DNA  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-4788

Query Match 11.6%; Score 113; DB 4; Length 1551;  
Best Local Similarity 53.0%; Pred. No. 6e-14;  
Matches 307; Conservative 0; Mismatches 245; Indels 27; Gaps 2;  
QY 53 CCTCCCTGCGGGGAGAGACCTTACCCGCTCTGCTCTCGAGGTCTTCTGC 112  
Db 560 CCTGCGCTGCAATGCGCCCAAGACGCTTCAAAATGCTCTCCGAACTGATGTCG 619  
QY 113 AGGAGACCGGGGTGAGAGACCGCTTATCCGCTTATCCGCTTCTGAGGCTTCCCA 172  
Db 620 AACCAACCGAGATACCAAGGATATCCCTTATTTGAACGCTTATAGCGGCTTCCCA 679  
QY 173 CCTGAAAGCCCTGCGCGCGCTTCCCTGGAAGAGTCTTAAAGGCTTGGAGGGGGCG 232  
Db 680 CGGTGTGATCTGCGCAAGCGCGCTGATGAAGTCTGATCTGATGAGACCGGTCTGG 739  
QY 233 GCTACTACCGCGCGCGCAAC-----CTCACCGCTGCGCGCGCAAGCG 277  
Db 740 GCTACTACCGCGCGCGCAACTTACATAAAGCGCGCAAGTCCCACTGCAAG 799  
QY 278 TGGAGAGCTTCCCGGAGCTTCCGCGAGCTTCCGCGGCTTCTGAGCTTACCA 337  
Db 800 GCGGGAATTCCTCCGAGCTTGAAGAGTGGCGCGCTACCGCGCTGCGCGCTCA 859  
QY 338 CCGCGCGCGCGCTGCTCAATCGCTTCCGCGAGCGGAGCGGCTGAGCGGGAACG 397  
Db 860 CCGCGCGCGCGATTTTATCTCTTCTGCTGCGTCAAGCATTTATCCGATTTCTGACGCGCAAG 919  
QY 398 TCCGAGGAGTCTTCTCCGCTCTTCCGCGGAAAGCCCAAGAGAGAGCTTTTCG 457  
Db 920 TGAAGCGGTCTGCGCGCGCTGCTATGCTATGAGCGGCTGCGCGGAAAAAAGAGTGG 979  
QY 458 CCTCGCCCAAGGCGCTTCCCGCGAGGCGTGAAGCGCGGCGGTGG-----A 505  
Db 980 AAAAAGCGCTGTGAGATCAAGAGAGGTACCCCGGAGAGGCGCTGAGCGCTTCA 1039  
QY 506 ACCAGGCGCTATGAGAGTCCGCGGCAAGCTTCTGCTGCGAAAGCGCGCTTTCGCGGG 565  
Db 1040 ACCAGGCGATATGATCTGCGGCGAATGTTTGAACCGCGCTGGAAGCGGAAGTGCAGC 1099  
QY 566 CTTGCGCGCTTCTGCGCGCGGCGGAGAGAGCGCC 604  
Db 1100 TGTGTCTGCTGAGCAACGCGTGTGCTGAGCGAACC 1138

RESULT 5  
US-09-252-991A-6783  
Sequence 6783, Application US/09252991A

Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6783  
LENGTH: 549  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6783

Query Match 11.0%; Score 107.2; DB 4; Length 549;  
Best Local Similarity 54.4%; Pred. No. 7.5e-13;  
Matches 248; Conservative 0; Mismatches 193; Indels 15; Gaps 1;  
QY 55 CTCCCTGCGGGGAGAGACCTTACCGGCTCTGCTCTCGAGTCTTCTGAG 114  
Db 60 CTGCGCTGCGAGAGGAGATCAACCTTACCGGCTGCTCTCGAGTCTTCTGAG 119  
QY 115 CAGACCGGAGTGAAGAGCGCTTCCCTATTAACCGGCTTCTGAGCGCTTCCAC 174  
Db 120 CAGACCGAGTCAAGACCGTGTGATTAACCGGCTTCTGAGCGGCTTCCAC 179  
QY 175 CTGAAGCGCTGCGCGGCTTCTCTGAGAGAGTCTTGAAGTCTTGAAGCGGCGGGC 234  
Db 180 GTGAGGCACTGCGCGGCGGCGGCGAGAGAGTCTTGAAGTCTTGAAGCGGCGGGC 239  
QY 235 TACTACCGGGGAGAGACCTTCAACCGCTGCGCGGAGAGCT-----G 279  
Db 240 TACTACCGGAGTGGCGGCAACCTTCAACCGGAGAGTCTTGAAGCGGAGAGCG 299  
QY 280 GAGAGCTTCCCGGAGCTTCCGCGAGCTTCCGCGGCTTCTGAGCTTCAAC 339  
Db 300 GGGAGTTCCTCCGCGAGCTGAGCAACTGCGGCACTGCGCGGAGTCTGAGCGCAAGC 359  
QY 340 GCGCGCGGCTGCTCTCATGCTTCTGCGGAGCGGAGCGGCTGAGCGGAGAGTCTC 399  
Db 360 GCTGAGCACTGCGCAACCTGCTGATGAGCGCTGCGGAGCGGAGTCTGAGCGCAAGTCTC 419  
QY 400 CGGAGGCTCTCTCCGCTTCTGCGCGGAGAGCGGAGAGAGAGAGAGCTTTTCCGC 459  
Db 420 AAGCGGATCTGCGCGCTTCTGCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGTCTC 479  
QY 460 CTCGCCAGGAGCTTCTCCCGAGAGGCGTGAAGCGG 495  
Db 480 AGGCGCTGTGGAAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 515

RESULT 6  
US-08-813-574-1  
Sequence 1, Application US/08813574  
Patent No. 6013473  
GENERAL INFORMATION:  
APPLICANT: Wei, Ying-Fei  
TITLE OF INVENTION: Human Muty  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

```

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/813.574
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/013,132
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1858 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-813-574-1

```

```

Query Match      10.7%; Score 104; DB 3; Length 1858;
Best Local Similarity 54.0%; Pred. No. 3.6e-12;
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

```

```

QY 69 GAGAGAGACCCCTTACCGCGCTCTGCTCCGAGGTCTTGTGACGAGACCGCGGTGGA 128
DB 498 GGAAGAGGCGGCAATATGCTGTGTGTCTGAGAGTATGTGTGACGAGACCGAGTTC 557
QY 129 GCAGGCGCTCCCTTATACCGCGCTTCTTGTGAGCGCTTCCACCTGAAAGCGCTGCG 188
DB 558 CACTGTATCAACTACTATACCGAGTGTGAGAGTGTGCTACCTGAGGAGCGCTGCG 617
QY 189 GCGGCTTCCCTGAGAGAGTCTTATGAGTCTGCGAGGCGGCGCTACACCGCGCGCG 248
DB 618 CAGTGCTTCCCTGAGAGAGTGTATCACTGTGCGCTGCGCTGCGCTGCTATTTCTGCG 677
QY 249 GGAACACTTCCACCGCTGCGCGAGCGTGTGAGAGCTTCCCGCGAGCTTC----- 300
DB 678 CCGCGCGCTGAGAGGAGAGCTCGAAGTGTGTAGAGACTAGGGGCGCAATGCGCAG 737
QY 301 -----GCGAGCTTGGGGGCTTCTGTGTCTGCGGCTTACACCGCGCGCGGT 350
DB 738 TACAGCAGAGACCTTGAGAGAGCTCTGCTGTGCGGTGCGCGCTACACAGCTGCGCAT 797
QY 351 GGCCTTCATGCGCTTCCGAGAGCGGCGTGTGAGCGGAAAGTCCGAGGGTCTCT 410
DB 798 TGCCTTATGCGCTTGTGCGAGGAGACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 857
QY 411 CTCCCGCTCTTGCGCC- GGAAGACCCCAAGAGAGAGAGCTTTTCCGCTGCGCCAG 468
DB 858 GTGCCGTGTCCGAGCATTTGTGTGTATCCACAGACCTTGTTCAGAGAGCTGTG 917
QY 469 GGCCTCTCTCCGAGGAGCGTGTGACCGG-----GGGGTGTGAAACAGGCGCTCAT 518
DB 918 GGGCTTAGCCACAGAGTGTGTGAGACCGCGCGGAGGAGATTTTCAACAGAGCGCAT 977
QY 519 GAGAGCTCGGGGCGACGCTGTGCTGCGGAAACGCGCGCTTGTGCGGGCGCTGCGCCCTAG 578
DB 978 GAGAGCTAGGGGCGACAGTGTGTATCCACAGCGCCCATGTGTGACGACGAGTCCCTGTGA 1037
QY 579 GGCCTTCTGCGCGG 592
DB 1038 GAGCCTGTGCGGG 1051

```

RESULT 7

```

US-09-651-656-38
; Sequence 38, Application US/09651656
; Patent No. 6340566
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
; TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,
; FILE REFERENCE: IL-10689
; CURRENT APPLICATION NUMBER: US/09/651,656
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 2293
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-651-656-38

```

```

Query Match      8.9%; Score 87.2; DB 4; Length 2293;
Best Local Similarity 54.1%; Pred. No. 7.2e-09;
Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;

```

```

QY 53 CCTTCCCTGCGCGGGGAGAGACCTTACCGCGCTCTGTCTCCGAGTCTTCTGCG 112
DB 1161 CTCTGCGCTGCGAATAATTGACAGACCGCTCAAGATATGCTCTCAGAGTATGTTGC 1220
QY 113 AGAGAGCGCGGTGAGAGAGCGCCCTCCCTATTACCGCGCTTGTGAGAGCGCTTCCCA 172
DB 1221 AACAACTAGAGTGTGACCGCTTATCCCTATTGAAAGCTTATATGCGCGCTTCCCGA 1280
QY 173 CCTGAGGCGCTGCGCGCGCTTCTCTGAGAGAGTCTTATGAGTCTGCGAGGCGCGG 232
DB 1281 CGGTAGCGATCTGCGCATGTGCGCGCTGCGAGAGTCTCCACTGTGTGAGACCGGCTTG 1340
QY 233 GCTACTACCGGCGGCGGAGAACCTTCCACCGCTGCGCGGAGCGTGTGAG----- 282
DB 1341 GCTATTACCGCGCGCGCGCATGTGCTAAGGCGGCAACAAGTGTGCGACTTACAG 1400
QY 283 -----GAGCTTCCCGAGCTTCCGCGAGCTTCCGCGGCGCTTCTGTGCGGCGCTTACA 337
DB 1401 GCGGTAAATTCGGAACCTTTGAGAGATTTGACAGACTCCGCGCGTCCGCGCTTCCA 1460
QY 338 CCGCGCGCGCGGTGCGCTTCCATCGCTTCTGCGAGCGGCGTGTGCGCGGTGTGAGCGGAG 397
DB 1461 CCGAGGCGCGATCTCTGCTTCTGTGAGTGTGAGACTTCCGATTTCTGAGCGGTAAAG 1520
QY 398 TCCGAGGAGTCTCTCCCGCGCTTTCG 425
DB 1521 TCAACGCGTGTGCGCGCTGTATGC 1548

```

```

RESULT 8
US-09-650-855-38
; Sequence 38, Application US/09650855
; Patent No. 6365355
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; TITLE OF INVENTION: LAWRENCE LIVERMORE NATIONAL LABORATORY
; TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA
; TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA
; FILE REFERENCE: IL-10284
; CURRENT APPLICATION NUMBER: US/09/650,855
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38

```

LENGTH: 2293  
TYPE: DNA  
ORGANISM: Escherichia coli  
US-09-650-655-38

Query Match 8.9%; Score 87.2; DB 4; Length 2293;  
Best Local Similarity 54.1%; Pred. No. 7.2e-09;  
Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;

53 CCCTCCCTGCGGGGAGAGACCTTACCGCGCTGTCTCCGAGGTCCTTCTGC 112  
DB CTCTGCTTGGCAATTGACAGACGCTTCAAGATGAGCTCTCAGAGTATGTC 1220  
QY AGCAGACCGGAGTGAAGAGACGCTTCCCTATTACCGCGCTTCTGAGCGCTTCCCA 172  
DB AACAACTCAGGTTGCGACCGTTATCCCTATTGTAACGCTTATGCGCGCTTCCCGA 1280  
QY CCTGGAAGCGCTTGGCGCGCTTCTCTGGAAGAGTCTTAAAGTCTGCGAGGGGCGG 232  
DB CGGTGACCGATCTCGCAATGCGCGCTCGACGAGTCTTCACTGTGAGACCGGCGCTTG 1340  
QY GCTACTACCGGCGGGAGAACACTCCACGCGCTGCGCGGAGCGTGAAG----- 282  
DB GCTATTACGCGCGCGCGCGCAATGCTGATTAAGGCGGACAAAGTGGCGACTTACAGC 1400  
QY 283 -----GAGCTTCCCGGAGCTTGGCGGAGCTTCTGTCTCGGCGCTTACA 337  
DB GCGGTAAATTCCCGGAAACCTTTGAGAGATGACACTGCGGCGCGCTGCTTCCA 1460  
QY 338 CCGGCGCGGCGGCTTCCATCGCTTGGGAGCGGCGGCGGCTGAGCGGAGC 397  
DB 1461 CCGGAGCGCGGATCTCTCGCTTCTCTGAGTGAAGCACTTTCGATTCGACGCTTACG 1520  
QY 398 TCCGAGGCTCTCTCCGCGCTTTCG 425  
DB 1521 TCAACCGGTGCTGCGCGCTGCTATG 1548

## RESULT 9

US-09-103-840A-2  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-2007.00  
CURRENT APPLICATION NUMBER: US/09/103.840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 8.9%; Score 86.6; DB 3; Length 4403765;  
Best Local Similarity 50.2%; Pred. No. 1.8e-08;  
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;  
QY 22 CTCCTGCGCTGTACCGGAAACGCGCGCTTCCCTGCGGCGG---GAGAGAGAC 78  
DB 4022631 CTCCTGCGCTGTATCAGAGATCGACCGGAGATCTGCGCTGCGAGACCGCGGTGTACG 4022690  
QY 79 CCTTACCGCGTCTGCTCGAGGTCTTCTGACGACAGACCGGCGTGAAGACGCGCTC 138

DB 4022691 CCGTGCACATCTGTATGACGAGTTCATGCTGACAGACACCGCGCGCGCGGTGCTG 4022750  
QY 139 CCTTATTACCGCGCTTCTTGTGAGCGCTTCCACCTTGAAGCGCTGCGCGGCTTTC 198  
DB 4022751 GCGATCTGCGCGGACTGCGGTGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 4022810  
QY 199 CTGGAAGAGTCTTATGAGTCTGCGAGGGGCGGCTTACCTGCGGCGGAGAACCTTC 4022870  
DB 4022811 ACCGCGAATGTGTACGCGCTGCGGAGAGTGGCTATCCAGGAGACAGCCCTTA 4022870  
QY 259 CACGCGCTGCGCGGAGCGTG-----GAGAGCTTCCCGGAGCTTTCGCC 303  
DB 4022871 CACGAGTGGCGCACCGCTTATCGCGCGGACACCAATGAGTGTGCTCCAGATATCGAG 4022930  
QY 304 GAGCTTGGGCGGCTTCTGTCTGCGGCTTACCGCGCGCGGCTGCTTCAATGCC 353  
DB 4022931 ATCCTGATCACCTGCGCGGCGCTGCGGAGCTACCGCGCGCGGCTGCTGCTTTCGCT 4022990  
QY 364 TTGCGGAGCGGCTGCGCGCGGCTGCGGAGAGAGTCTTCTGCGCGCTTTC 422  
DB 4022991 TACCGCGAGCGGCTGCGGTGCGGACCAATGTCGCGGCTGCTGCGCGCGCTT 4023050  
QY 423 -----CGCGCGGAAAGCGCGGAGAGAGCTTTCGCGCGCTTTCGCGCGCG 471  
DB 4023051 CACGCGCGCGCGCACCGCGGCTGCGCATGCTGCGCGCGCGCGCGCGCGCTTGGCG 4023110  
QY 472 CTCCTCCCGAGGCGGTGACCGCGGCGGTGAGACCAAGCGCTTATGAGTCTGCGGCG 531  
DB 4023111 CTCCTGCGCGACCGCGAGACCGCGGCTGATTTTCGCTGCGCTGATGAGTGGGTGCG 4023170  
QY 532 ACCGTCTGCTGCGCGGAAAGCGCGCGTGTGCGGCGCGCTTTCGCGCGCGCTTTCGCGCG 591  
DB 4023171 ACAGTGTGACCGCGCGCACCGCGCGCGGCTGAGTGTGCGCGCGCTGAGCTGCGCATGG 4023230  
QY 592 GGGAGGAGGCGCGCGCGGCTTACCGCGCGCGGAGAGCGCGGCGGAGAGAGCGC 651  
DB 4023231 CCGCATGCGCGTATCCGCGCTGCGAGCGTCC---GCGCGCGCGGCGGCGCGCTACAC 4023287  
QY 652 CTCGTGCGCTGCTCTCTCTGCGGCGG 678  
DB 4023288 GGAACGACCGCGCAAGTCCGCGAGCG 4023314

## RESULT 10

US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-2007.00  
CURRENT APPLICATION NUMBER: US/09/103.840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 8.9%; Score 86.6; DB 3; Length 4411529;  
Best Local Similarity 50.2%; Pred. No. 1.8e-08;  
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;  
QY 22 CTCCTGCGCTGTACCGGAAACGCGCGCTTCCCTGCGGCGG---GAGAGAGAC 78

```

Db 4030550 CTTCGCTTGTGATCAGGATGCAACCGGATCTGCTGCGAGAGCCCGGTGTCAAC 4030609
QY 79 CCTTACCGGCTCTGCTGCTTCCAGGCTCTTCTGACAGAACCCGGGTGAGAGCCCTC 138
Db 4030610 CCGTGGAGATCTGCTGACGCGATTCATGCTGACACACCCGCGCCGCGGGGTGCTG 4030669
QY 139 CCCTATTACCGCGCTTTCTGAGCGCTTCCACCTGAAAGCCCTGCGCCGCTTCC 198
Db 4030670 GCGATCTGGCGGACCTGGGTGGCGGCTGGCCACCGCTCGACCAACCGCGAGC 4030729
QY 199 CTGGAAGAGGTCTTAAAGGTCTGGACAGGGGCGGGCTACTACCGCGGCGAGAACCTTC 258
Db 4030730 ACCGCGGATGTGTACCGCTGAGGCAAGCTGGGTATCCAGGCGAGCCAGGCTTA 4030789
QY 259 CACCGCTGCGCGGAAAGCGTG-----GAGAGACTTCCCGGAGCTTGGCC 303
Db 4030790 CACGAGTGGCGCAACCTGATGCGCCGCGACCAATGACGTGTGCTCCGAGATATCGAG 4030849
QY 304 GAGCTTGGGGGCTTCTGCTGCTGAGGCTTACACCGCGCGCGCGGCTGCTCATCGCC 363
Db 4030850 ATCTGTGATCACCCTGCGGCGCTGCGGAGCTACACCGCGCGCGGCTGCTGCT 4030909
QY 364 TTGCGGGAGCGGGTGGCGGGTGGACGGAACGTCCGAGAGGTCTCTCCGCTCTT- 422
Db 4030910 TACCGCCAGCGGGTCCGGTGTGACCAATGTGCGCGCGGTGTGCGCGCGCTT 4030969
QY 423 -----CGCCGGGAAAGCCCAAGAGAGAGCTTTTCGCTGCGCCGAGG 471
Db 4030970 CACGCGCGCGCGACGCGCGGTGCGCATGCGTGCCTCCGACACAGCGAGCTTTGGCG 4031029
QY 472 CTCCTCCCGAGAGGGTGAACCGGGGGTGTGAACCAAGGCTCTATGAGCTCGGGCC 531
Db 4031030 CTGTTCGCGACCGGAGACCGCGCTGAATTTTCGCTGCGCTGATGAGATTGGTGGCG 4031089
QY 532 ACGGTCTGCTGCGGAAAGCGCGCGGTGGCGGCTGCGCGCTGAGGAGGCTTCTGCGCG 591
Db 4031090 ACGGTGTGACCGCGCGCGACACCGCGGTGTTATGCGCTGAGCTGTGGATGG 4031149
QY 592 GGGAAAGAGCGCGCGCGGCTTACCCCGCGCGCGCGCGAGAGCGCGAGAGAGAGCGC 651
Db 4031150 CGGATGCGCGGTATTCGCGCTCGACGAGTCC--GCCGCGCGGGGCGAGGCTTACAC 4031206
QY 652 CTCGTGCGCTGCTGCTGCTGCGGGCG 678
Db 4031207 GGAACCGACCGCAAGTCCGCGAGCG 4031233

```

```

RESULT 11
US-08-961-527-12/c
; Sequence 12, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunesh
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8504
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-12

```

```

Query Match      8.4%; Score 82; DB 4; Length 9909;
Best Local Similarity 52.2%; Pred. No. 8.5e-08;
Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

```

```

QY 22 CTCCTGCTGCTGATACCGGAAACCGCCGCTCCCTGCGCGGGGAGAGAGCCCT 81
Db 5808 CTTCGCTTGTATGATGAACAAAGAGATTGCTTGAGAGAGAAATTAATCT 5749
QY 82 TACCGGCTGTGTCTCCGAGGTCTTTCGACAGACCGGGTGGACAGGCCCTCCC 141
Db 5748 TATCATCTGGATCTGAATCATGCTTCAGACAGACAGGGTGTATACATTATCCT 5689
QY 142 TATTACCGCGCTTCTGAGAGCGCTTTCACACCGTGAAGGCCCTGCGCGGCTTCCCTG 201
Db 5688 TACTACGAAAGATTTTTGACTGATGCTGTTTCAACTGTGCAAAATCTGGCACTGGCCTGAG 5629
QY 202 GAAAGGCTCTTGAAGGTGTGCGAGGGGCGGCTA-----CTACCGGCG 246
Db 5628 GAGAGTCTTGAAGAGCTTGGAGGGCTTGAGCTTATATTCTGAGTTGCCAATATGACG 5569
QY 247 GCGGAACACTTCAACCGCTGCGCGGAGAGGTGAGAGAGTTCCTCCGAGCTTCCGCGAG 306
Db 5568 GCTGAGCGCCAGCATTAATTAAGTACTGACTTGTGGCCCAATTTCCAAATACCTTGAAGA 5509
QY 307 CTTCGGGGGCTTCTGCTGCTGCGGCTTACACCGCGCGCGGTGAGCTTCATGCGCTTC 366
Db 5508 ATTTCAGCTTGAAGAGGATTTGACCTTACAGACGAGAGCCATTTCAGATATGCTTTT 5449
QY 367 GGGAGGGGGTGGGGGGGTGAGCGGAGACGTCCGAGAGGTCTCTCCGCTCTTTCG 424
Db 5448 AACTTGCTGAGCCAGCTGTAGATGTATGATGATGCGGGTCTTGGCGGCTGTGTTG 5391

```

```

RESULT 12
US-09-107-532A-3654
; Sequence 3654, Application US/09107532A
; Patent No. 6581275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998

```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 893-5007  
TELEFAX: (781) 893-8277  
INFORMATION FOR SEQ ID NO: 3654:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1212 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...1212  
SEQUENCE DESCRIPTION: SEQ ID NO: 3654:  
US-09-107-532A-3654

Query Match  
Best Local Similarity 47.6%; Pred. No. 0.00029; Indels 27; Gaps 2;  
Matches 278; Conservative 0; Mismatches 279;

31 TGGTACCGGAAAAAGCCGCCCTCCCTGCGCGGGGAGAGACCTTACCGGCTC 90  
100 TGGTATACAGAAAAAGAAACCTGCTGGCGCTATATCGTATCTTATCGTATC 159  
91 CTGGTCTCGAGGTCTCTTTCGACAGACCCGGGTGAGACGCGCTTCCCTTATCCGC 150  
160 TGGATCTCTGAGATCATGCTTTCAGCAACAGAGTACACAGATGATGATTTAT 219  
151 CGCTTTTGAGAGCCTTTCACCTGAGGCGCTGCGCGGCTTCCCTGAGAGAGTCT 210  
220 CGTTTCAAGAAATGCTTCTCAATTAAGAGCTTACAAATGCTTCAAGAAAGAACTT 279  
211 CTGAGGTCTGACGCGGCGGCTACTACCGCGCGGCAACCTC-----258  
280 TTTAAACATGAGGAGGCGCTTGTATTAATCAAGAGCGCGAATCAAGCTGCTGCT 339  
259 ---CACCCTGAGCCGAAAGGTGAGAGAGCTTCCCGAGCTTCCCGAGCTTCCGAGG 315  
340 AATAGATCATGTGAGAAATTTGATGGGGAATGCGCAACGCTGAAGAAATTAAGCTCA 399  
316 CTTCCTGCTCTCGGCGCTTACACCGCGCGCGGCTGCTTCCCTTCCGAGAGCGG 375  
400 TTGAAGGAAATGAGACCATATACAGAGCGCGATGGAATGCAATTTGAGCTTCA 459  
376 GTGCGGCGGTGAGACGGAACGTCGAGAGGCTCTTCCCGCTCTTTCGCGCGGAAAGC 435  
460 GAACCTGAGTTGACGCGACGTGATGAGAGTACAGAGATTTGTTGATTTGAAAGCA 519  
436 CCCAAGAGAA-----GAGCTTTTCCCGCTCGCGCAAGGCGCTTCCCGAG 483  
520 GACATCGCAAAAGCTTCTAGCAGAGAAATCTTTGATGAGAGGATGCGGAAATCATTTGAT 579  
484 GCGGTGACCCCGGGGTGTGAAACAGGCGCTCATGAGCTTCCGCGGCAAGCTGCTGCTG 543  
580 GAAAGCATTCAGGATGAAATTAACCAAGCAATGATGATCTAAGGTTGACGCTATCTGATCA 639  
544 CCGAAACGCGCCCGTTGCGGGGCTGCGCCCTAAGGAGCTTCTG 587  
640 CCAACTTCTCCAAAGTGTGAGACTTGTGCTGATCCAAAGCTTTTG 683

RESULT 13  
US-09-134-000C-1710  
Sequence 1710, Application US/09134000C  
Patent No. 6617156  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-032  
CURRENT APPLICATION NUMBER: US/09/134,000C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/055,778  
PRIOR FILING DATE: 1997-08-15  
NUMBER OF SEQ ID NOS: 6812  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1710  
LENGTH: 1188  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
US-09-134-000C-1710

Query Match  
Best Local Similarity 48.8%; Pred. No. 0.0005; Indels 15; Gaps 1;  
Matches 211; Conservative 0; Mismatches 206;

8 CTGCGGGAAGCCCTCTCTGCTGATCCGGAAGAACCGCCCTCCCTGCGGCGG 67  
53 CATTCAGGAAGATTTTATGCTTGTATGACGAGAAAAAGCAATTTACCTTGGCGAG 112  
68 GGGAGAAGACCTTACCGGCTGCTGCTCCGAGGTCTCTGACAGACAGCCGGGTG 127  
113 CGAATACAGATGATATTCGATTTGATTTCTGAATTTATGCTACAAACCTCGGTAG 172  
128 AGCAGGCGCTCCCTATTTACCGCGCTTCTGAGAGCGCTTCCACCTGAGAGCGCTG 187  
173 ATACAGTATGATTTATTTATGATTTATGATTTATGATTTATGATTTATGATTTATG 232  
188 CCGCGGCTTCCCTGGAAGAGTCTTATGAGTCTGCGAGGCGGCGCTACTACCGCGG 247  
233 CGAAGGCGCAGATGATGATTTGATTAAGCTTGGAGAGGTTAGTTACTATTTACAGAG 292  
248 C-----GAAACACTCCACCGCTGCGCGCAAGCTGAGAGACTTCC 292  
293 CGGCTAATTTAAAGTGGACGCGCAACAGATTTGTTCAAGATTTGGGAAAAATGCTG 352  
293 CGAGCTTCCGAGCTTCCGAGGCTTCTGCTGCGGCTTACACCGCGCGCGGTG 352  
353 ACACATCGAAGATTTGAGATTTAAAGAAATCGGCGCTGATGCGATTTG 412  
353 CTTCATCGCTTCCGAGAGCGGCTGCGGCGGTGAGAGGAACTTCCGAGAGGTCTCT 412  
413 GCAGTATTTGCTTAAATCTCCGGAACAGCAATTAACGATTAACGATGCGGATGTA 472  
413 CCGGCTCTTCTG 424  
473 GCGGTTATTTG 484

RESULT 14  
US-09-385-028-23  
Sequence 23, Application US/09385028  
Patent No. 6232106  
GENERAL INFORMATION:  
APPLICANT: Susan E. Jensen  
APPLICANT: Kwame A. Aido  
APPLICANT: Ashish S. Parthasarathy  
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
Patent No. 6232106  
TITLE OF INVENTION: Acid Biosynthesis  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC

STREET: The Jennifer Building, 400 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/385,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/790,462  
FILING DATE: 29-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: D. Douglas Price  
REGISTRATION NUMBER: 24,514  
REFERENCE/DOCKET NUMBER: 1418/P57452US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-6666  
TELEFAX: (202) 39305350  
TELEX: RCA 248593 IDEA UR  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1227 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-385-028-23

Query Match 6.4%; Score 62.2; DB 3; Length 1227;  
Best Local Similarity 44.4%; Pred. No. 0.00055;  
Matches 345; Conservative 0; Mismatches 423; Indels 9; Gaps 2;

157 CTGGAGCGCTTCCACCTGAGAGCCCTGCGCGGCTTCCCTGAGAGGCTCTTAGG 216  
88 CTGGCGGGGCTGCGGCTCCAGAGGCGCGAGCGGGTGAAGCTGTGGGAGCGGAGCAG 147  
217 GTCTGCGAGGCGGGGCTATACCGCGCGGCGGAGACACCTCCACCGCGCTGGCGGAGC 276  
148 GTGTGGCTGTGACCTGCGACGCGGGGCGCGGCGGCTGTGGGCGAGCCGCCGTTTAC 207  
277 GTGAGAGAGCTTCCCGAGCTTCCGCGAGCTTCCGCGGCGCTTCTGTGCTCGGCGCTTAC 336  
208 GCGGTGACGAGCGCGCGGCTTCCGATGCTGACCGGACCTCCAACTGTGTGCGGCGC 267  
337 ACCGCGGCGGCGGCTTCCATGCGCTTCCGCGGAGCGGGGTGGCGGCGGTGAGCGG--- 392  
268 AACCGGAGTGGGCTGCTTATCGCATGAGACGACCGCGAGCACTCCCGGCTGCGTGC 327  
393 --GAACGTCGCGAGGCTCTCCCGGCTTCCGCGGCGGAGAGCCCGCAAGAGAGAG 450  
328 ATGTACACCGGAGCTTCTGGCGCGCGCGCGCGCGAGGCGCTGCGCGCGCGGTGCGGAG 387  
451 CTTTTCGCTCGCGCGAGGCTCTCTCCCGAGGCGGTGAGACCGCGGGGTGTGAGACGAG 510  
388 CTGCTGAGAGAGATCTGGGGGCGGTGTGAGAGGGGAGAGCGGCGGTGACCTGTGCGCC 447  
511 GCGCTCATGAGAGCTGGGGGCGAGGCTGTGCTGCGCGAGAACGCGCCCGCTTGGGGGCGTGC 570  
448 GAGCTGAGAGATCCCGGTGCTCTCGCGGATCACTGCTTTCGGGCGCGGTGAGACGAC 507  
571 CCCCTAGAGGCTTCTGCGCGGAGAGAGCGCGCGCGCGCTACCCCGCGCGAGAGAG 630  
508 GCGCGGAGGTTATGAGAGAGCGGAGCGCGGCTGCTCATGAGACCGGCGTTACACCGCGAG 567  
631 CGCCGAGGAGAGAGAGCGCTGCTGCGCGCTGCTGCTGCTGCGGCGAGAGGGGTGAC 690  
568 CAGGTGCGCAAGGCGCGGAGAGAGCTGAGCGGCTATCTGGGAGAGCTGTGTGAGAGAGCGG 627

691 CTGAGAGGCTTGAAGGCGCGCTTCCAGAGGCTCTACGCGCTCCCGCTTCTTCCCGCTGAG 750  
628 ATCGAGAACCGCGGACCGACCGATGATCGCGCGCTCTCATGACAGGAGGCGCGGAG 687  
751 GAGCTTCCCGGCGGAGGCGGCGCTTGGGGGTGAGGTCTAGGCGCCCTAGGCGAGTGC 810  
688 CATCTGCGGCTGAGAGAGATGCTCCATGTGCGGCTGCTGCTGTGCGCGGCTGACGCG 747  
811 CAGCGCTTACCGACCGGAGGCTTCCGCTGAGAGGTGCGGCGGCGCTTGGAGAGGAG 870  
748 ACACACACCGAGGAGGAGCTGAGCTGCTGAGCTGCTACCGGACCGGAGCTGCGC 807  
871 GGGAGAGACCTTGAAGAGGCGC---CTACCGCACTCATGAGAGAGTGTCCGC 924  
808 GGGCGCTTACCGAGGAGCGCGCGCTGCTGCGCAAGGCGGTGAGAGAGCTGCTGCGC 864

RESULT 15  
US-09-726-614-23  
Sequence 23, Application US/09726614  
Patent No. 6514735

GENERAL INFORMATION:  
APPLICANT: Susan E. Jensen  
APPLICANT: Kwamena A. Aidoo  
APPLICANT: Ashish S. Paraskar  
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
Patent No. 6514735  
TITLE OF INVENTION: Acid Biosynthesis  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSER: JACOBSON, PRICE, HOLMAN & STERN, PLLC  
STREET: The Jennifer Building, 400 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/726,614  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/790,462  
FILING DATE: 29-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: D. Douglas Price  
REGISTRATION NUMBER: 24,514  
REFERENCE/DOCKET NUMBER: 1418/P57452US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-6666  
TELEFAX: (202) 39305350  
TELEX: RCA 248593 IDEA UR  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1227 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-726-614-23

Query Match 6.4%; Score 62.2; DB 4; Length 1227;  
Best Local Similarity 44.4%; Pred. No. 0.00055;  
Matches 345; Conservative 0; Mismatches 423; Indels 9; Gaps 2;

157 CTGGAGCGCTTCCACCTGAGAGCCCTGCGCGGCTTCCCTGAGAGGCTCTTAGG 216  
88 CTGGCGGGGCTGCGGCTCCAGAGGCGCGAGCGGGTGAAGCTGTGGGAGCGGAGCGGAG 147



QY 217 GTCTGCGAGGAGGCGGCTACTACCGGCGGCGGAGACCTTCCACCGCGCTGGCCCGAGC 276  
 Db 148 GTGTGCTGTGTGACCTGCGACGCGGCGGCGCGGCGCGCTTCTGCGCGACCGCGCTTCACC 207  
 QY 277 GTGAGAGAGCTTCCCCCGAGCTTTCGCGAGCTTTCGCGGCGCTTCTGCTGCGGCTTAC 336  
 Db 208 GCGGTGACGAGCGCGCGGCTTTCGAGTGTACCGGACCTTCCAACTGTGTGCGCGC 267  
 QY 337 ACCGCGGCGGCGGTGCGCTTCATCGCTTCGCGGAGCGGCTGCGCGCGGTGAGCG---- 392  
 Db 268 AACCGGAGTGTGCGGTGTTCATTCGATGAGAGACCGGAGCACTCCGCGCTGCGCTCG 327  
 QY 393 --GAACTTCGAGAGGCTCTCTCCGCTCTTTCGCGGAGAACCCCAAGAGAGAG 450  
 Db 328 ATGCTCACCGGAGCTTCTGCGCGCGCGCGCGCGCTGCGCGCGCGCGGTGCGGAG 387  
 QY 451 CTTTTCGCGCGCGCGCGCTCTCCGCGGCGGTGAGACCGGCGGCGGTGAGACGAG 510  
 Db 388 CTGCTGAGAGAGATCTGAGGCGGCTGTGAGAGGAGAGCGGCGGTGAGCTGTGCGC 447  
 QY 511 GCCCTCATGAGCTTCGCGGCGGCGGCTGCTGCGGAGACGCGCGCTTTCGCGGCGCTGC 570  
 Db 448 GGAATTGACGATCCCGGTGCTTCGCGGGTATGACCTGCTTTCGCGCGCGGTGAGAGAC 507  
 QY 571 CCCCTAGGCGCTTTCGCGGCGGAGAGAGCGCGCGGCGCTACCCCGCGCGAGAG 630  
 Db 508 CGCGGAGGATTCATGAGAGACCGGAGCGGCTCTCATGACCGCGCTACACCGCGAG 567  
 QY 631 CGCGGCGGAGAGAGAGCGCTGCTGCGCTCTGCTCTGCGGCGGAGAGGAGGTGAC 690  
 Db 568 CAGGTCCGCAAGCGCGGAGAGAACTGACGCTATCTGCGGAGCTGTGAGAGAGGCG 627  
 QY 691 CTGGAAGGCTTGAAGGCGGCTTCAGGCGCTTACGCGCTTCCCTTTCCTGAG 750  
 Db 628 ATCGAGAACCGGAGACCGGACTGATGACGCGGCTGTGATGACAGAGTGGCGGCGG 687  
 QY 751 GAGCTTTCGCGGCGGAGAGCGGCTTTCGCGGCTGAGGCTTACGCGCTTACGAGTGGC 810  
 Db 688 CATCTGCGGCTGAGAGAGATGTCCGATGTGCGGCTGTGCTGTGTGCGGTACGCGC 747  
 QY 811 CACGCGCTCACCCGAGCGGAGCTTTCGCTGAGAGTGTGCGGCGGCGCTTTCGGAAGGAG 870  
 Db 748 ACCACCAACAGCGGAGCGGCTGAGCTGTGCTGAGCTGTACCGAGCCGAGCTGGCC 807  
 QY 871 GGGGAGAGCCCTGGAAGAGGCGC---CTACCAAGCTCATGAGAGAGTGTGCGC 924  
 Db 808 GGGCGCTCATCGAGAGACCGGCGCTGTGCTCCAAAGCGGTGAGAGAGTGTGCGC 864

Search completed: October 8, 2004, 05:13:15  
 Job time : 114 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 13:53:05 : Search time 66 Seconds  
(without alignments)  
718.508 Million cell updates/sec

Title: US-09-938-901-2

Perfect score: 1700

Sequence: 1 VEAMRKALLAWYRENNRPLP.....VLRKALPLLAHGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 541936

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1700	100.0	325	11 US-09-938-901-2	Sequence 2, Appl1
2	486.5	28.6	486	9 US-09-925-301-1326	Sequence 1326, Ap
3	377	22.2	313	15 US-10-156-761-12241	Sequence 12241, A
4	365	21.5	293	10 US-09-738-626-6433	Sequence 6433, Ap
5	199	11.7	268	10 US-09-864-866-443	Sequence 43, Appl
6	184.5	10.9	310	15 US-10-156-761-12127	Sequence 12127, A
7	147.5	8.7	260	10 US-09-738-626-3828	Sequence 3828, Ap
8	145	8.5	211	9 US-09-912-020-296	Sequence 296, App
9	128	7.5	281	15 US-10-128-714-3128	Sequence 3128, Ap
10	128	7.5	461	15 US-10-128-714-8128	Sequence 8128, Ap
11	112.5	6.6	1332	12 US-09-840-743-11	Sequence 11, Appl
12	111	6.5	1729	12 US-09-840-743-2	Sequence 2, Appl1
13	110	6.5	1413	12 US-09-840-743-8	Sequence 8, Appl1
14	108.5	6.4	635	12 US-10-302-840A-6	Sequence 6, Appl1
15	104.5	6.1	549	15 US-10-156-761-14029	Sequence 14029, A

16	103.5	6.1	298	12 US-09-840-743-33	Sequence 33, Appl
17	102	6.0	1133	15 US-10-156-761-11645	Sequence 11645, A
18	100	5.9	807	15 US-10-044-692-5	Sequence 5, Appl1
19	100	5.9	807	15 US-10-044-539-5	Sequence 5, Appl1
20	100	5.9	1132	10 US-09-990-080-2	Sequence 2, Appl1
21	100	5.9	1132	10 US-09-749-728B-31	Sequence 31, Appl
22	100	5.9	1132	10 US-09-843-676-225	Sequence 225, App
23	100	5.9	1132	10 US-09-853-052-2	Sequence 2, Appl1
24	100	5.9	1132	12 US-10-295-681-57	Sequence 57, Appl
25	100	5.9	1132	15 US-10-053-758-225	Sequence 225, App
26	100	5.9	1132	15 US-10-208-243-2	Sequence 2, Appl1
27	100	5.9	1132	15 US-10-054-295-225	Sequence 225, App
28	100	5.9	1132	15 US-10-054-611-225	Sequence 225, App
29	100	5.9	1132	15 US-10-105-963-2	Sequence 2, Appl1
30	100	5.9	1132	15 US-10-044-692-2	Sequence 2, Appl1
31	100	5.9	1132	15 US-10-044-539-2	Sequence 2, Appl1
32	100	5.9	1154	15 US-10-044-692-323	Sequence 323, App
33	100	5.9	1154	15 US-10-044-539-323	Sequence 323, App
34	100	5.9	1189	15 US-10-044-692-325	Sequence 325, App
35	100	5.9	1189	15 US-10-044-539-325	Sequence 325, App
36	100	5.9	1200	15 US-10-044-692-324	Sequence 324, App
37	100	5.9	1200	15 US-10-044-539-324	Sequence 324, App
38	100	5.9	1285	15 US-10-044-692-314	Sequence 314, App
39	100	5.9	1285	15 US-10-044-539-314	Sequence 314, App
40	100	5.9	1407	15 US-10-044-692-334	Sequence 334, App
41	100	5.9	1407	15 US-10-044-539-334	Sequence 334, App
42	98	5.8	748	11 US-09-880-505-154	Sequence 154, App
43	98	5.8	748	14 US-10-051-643-154	Sequence 154, App
44	98	5.8	842	9 US-09-815-242-11950	Sequence 11950, A
45	97.5	5.7	412	15 US-10-156-761-8018	Sequence 8018, Ap

# ALIGNMENTS

RESULT 1  
US-09-938-901-2  
; Sequence 2, Application US/09938901  
; Publication No. US20030008291A1  
; GENERAL INFORMATION:  
; APPLICANT: Kuramoto Sei-ki,  
; APPLICANT: Yokoyama Shigeyuki  
; TITLE OF INVENTION: GENE ENCODING DNA REPAIR ENZYME  
; FILE REFERENCE: PH-1261-US  
; CURRENT APPLICATION NUMBER: US/09/938, 901  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: JP2001-47762  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Thermus thermophilus  
US-09-938-901-2

Query Match	100.0%	Score 1700;	DB 11;	Length 325;
Best Local Similarity	100.0%	Pred. No. 2,6e-152;		
Matches 325;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	VEAMRKALLAWYRENNRPLPWRGKEDPYRYLVSEVLLQOTRVGOALPYRRPLERPTLK	60	
DB	1	VEAMRKALLAWYRENNRPLPWRGKEDPYRYLVSEVLLQOTRVGOALPYRRPLERPTLK	60	
QY	61	ALAAASLEEVYRWOGAGYVRAEHLRLARSVEELPPSFELRGJLGLGPTAAVAST	120	
DB	61	ALAAASLEEVYRWOGAGYVRAEHLRLARSVEELPPSFELRGJLGLGPTAAVAST	120	
QY	121	AFGEVAAVADGNVRRVLSRLFAESPEKELFLAAGCLLEPGVDPGVWNAQALMELGATVC	180	
DB	121	AFGEVAAVADGNVRRVLSRLFAESPEKELFLAAGCLLEPGVDPGVWNAQALMELGATVC	180	
QY	181	LPRRPGACPLGAFRCGRKAPRRYAPARRRAKKEETVALVLLGKRGVLELDEGRFQG	240	

```
Db 181 LPRKRCGACPLGAFCCGKEAPGRYPAPRRKRAKEERLVALVLLGRKGVHLERLEGRRG 240
Qy 241 LYGVPLEPPELPGRREAFVGRSRPLGEVRLTHRLRVVRGALWEGEEDPWKRPLP 300
Db 241 LYGVPLEPPELPGRREAFVGRSRPLGEVRLTHRLRVVRGALWEGEEDPWKRPLP 300
Qy 301 KLMKVLKRALPLLAHAGVPLPDA 325
Db 301 KLMKVLKRALPLLAHAGVPLPDA 325

RESULT 2
US-09-925-301-1326
; Sequence 1326, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/0500/05862
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1326
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (438)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (447)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1326

Query Match 28.6%; Score 486.5; DB 9; Length 486;
Best Local Similarity 36.6%; Pred. No. 1.9e-37;
Matches 136; Conservative 46; Mismatches 107; Indels 83; Gaps 13;

Qy 1 VEAMRKALLAWYRENAARPLPWRG-----EKDPYRVLYSEVLLQOTFVEQALFYRRF 52
Db 104 VTAFFGSLISWYDQERKRLPMRRRAEDMDLDRAYAVAVSEVMLQOTQVATVINYTGW 163
Qy 53 LERFPTLALAASLEEVLRVWOGAGYRRAEHLRLARS-VEEL---PPSFAEL-RGL 106
Db 164 MOKWPTLQDASLASLEEVQMLMAGLGYSRGRRLDGGAKYVEELGGMPTAEFLQQL 223
Qy 107 PELGPTYAAVAASIAFERVAADVGNVRRLSLRFA----RESPEKEFLAAGLPE 161
Db 224 PCVGRRTAGATASIAFGATGVVDGNVAVLRCVRAIGADPSSTLVSQLMGLAQOLV-D 282
Qy 162 GVDPCGVWNOALMELGATVCLPKRRPGACPLGAFCRGKE----- 200
Db 283 PARPGDFNOAMMELGATVCTPQRPLCSQCPVESLCRARGRVBOEDLLASGLSGSPDVEE 342
Qy 201 -RP-----GRYPAPRRKRAK---EERLVALV-----LGRKGVHLE 232
Db 343 CAPNTGQCHLCLPSPRPMDTLGAVNFPKASRKRPPRESSSMTCVLEOGALGAQOLLVQ 402
Qy 233 R-LEGFRGGLYGP--LPPPEELPGRRAA-----GVSRPLGEVRLTHRL 278
Db 403 RNSGLLAGLWEPSPVTVWEPSEQLKALLQELQRYAGRPATPHKHLGEVYVTHSHKL 462
Qy 279 RVEVGRALWEGE 290
```

```
Db 463 TYGYVGLALEGQ 474

RESULT 3
US-10-156-761-12241
; Sequence 12241, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12241
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12241

Query Match 22.2%; Score 377; DB 15; Length 313;
Best Local Similarity 39.6%; Pred. No. 2.3e-27;
Matches 97; Conservative 32; Mismatches 88; Indels 28; Gaps 7;

Qy 2 EAMRKALLAWYRENAARPLPWR-RGEKDPYRVLYSEVLLQOTFVEQALFYRRFLEPFTLK 60
Db 25 EALHNOYIAWFDENARHRLDPMRRPDAGPWGVSEFMLOQTFVNVLLVEYEQMLRWPRPA 84
Qy 61 ALAASLEEVLRVWOGAGYRRAEHLRLARSVEE---LPPSFAELRGLPGIGPTTAA 115
Db 85 DLAKRPAEAVRANGRLGYPRRALRLHGAVALTERHNGDVPTEHAOLLAIPGIGEYTA 144
Qy 116 AVASTIAFERVAADVGNVRRLSLRFA-----ARSPKEKEFLAAGLPGDVPGV 167
Db 145 AVASFAVQORAAVLDITNVRVFAVAVTGQYTPMATTAAEKK-----LARALLPDESTAS 200
Qy 168 -WNOALMELGATVCLPKRRPGACPLGAFCCGKEA--PGRYPAPRK-----BRAKEER 217
Db 201 RMAAASMLGALVCTANKNETCHRCPLAGQCAMRLAGKPEHNDGPPRRGQTVAGTORVGR 260
Qy 218 LVALV 222
Db 261 LLAVL 265

RESULT 4
US-09-738-626-6433
; Sequence 6433, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
```

```

:
: CURRENT APPLICATION NUMBER: US/09/7738, 6526
:
: CURRENT FILING DATE: 2000-12-10
:
: PRIOR APPLICATION NUMBER: JP 99/377484
:
: PRIOR FILING DATE: 1999-12-16
:
: PRIOR APPLICATION NUMBER: JP 00/159162
:
: PRIOR FILING DATE: 2000-04-07
:
: PRIOR APPLICATION NUMBER: JP 00/280988
:
: PRIOR FILING DATE: 2000-08-03
:
: NUMBER OF SEQ ID NOS: 7059
:
: SOFTWARE: PatentIn ver. 3.0
:
: SEQ ID NO 6433
:
: LENGTH: 293
:
: TYPE: prt
:
: ORGANISM: Corynebacterium glutamicum
:
: US-09-738-626-6433

```

Query Match	21.5%	Score 365;	DB 10;	Length 293;
Best Local Similarly	34.68;	Pred. No. 2.9e 26;		
Matches 99;	Conservative 47;	Mismatches 120;	Indels 20;	Gaps 7

[illegible]

```

RESULT 5
US-09-864-866-43
: Sequence 43, Application US/09864866
: Patent No. US2002012756A1
: GENERAL INFORMATION:
: APPLICANT: Lloyd, R. Stephen
: APPLICANT: McCullough, Amanda K.
: APPLICANT: Nguyen, Khoa
: TITLE OF INVENTION: DNA REPAIR POLYPEPTIDES AND METHODS OF USE
: FILE REFERENCE: 365.00170101
: CURRENT APPLICATION NUMBER: US/09/864,866
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/206,279
: PRIOR FILING DATE: 2000-05-23
: NUMBER OF SEQ ID NOS: 49
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 43
: LENGTH: 268
: TYPE: PRT
: ORGANISM: Micrococcus luteus
: US-09-864-866-43

```

Query Match	11.7%	Score 199;	DB 10;	Length 268;
Best Local Similarity	32.4%	Pred. No. 1.2e-10;		
Matches 70;	Conservative 31;	Mismatches 95;	Indels 20;	Gaps 8

0y 27 PYVALSELLDQ---RVEQALPYRRFLEFPFLKLAASAEEVLRVAWGAGYRR-R 82  
 : :::: | | | | | | | | | | | | | | | | | |  
 Db 44 PELLATATLSAQTDVRVNAATP--ALFAFRPDHAMAATAPELLOELVRTGYFRNK 100  
 : : | | | | | | | | | | | | | | | | | |  
 0y 83 AEHLRLARSV-----EELPRSFALLRCLPGICGPTAAVAASIAFGERVAAVDGNRVVL 137  
 | : | | | | | | | | | | | | | | | | | :

Db 101 ASAILRLSGELVGRHGDGEV/PARLEEDLVALPGVGKRTAEVVLGNMFGQDGITVDTHFGRLA 160

Qy 138 SRL-FARESEKKEKELFALAQGLLPESDYGVMNQALMELGATVYLLPFRRCGACPLAFC 166

Db 161 RRLGTFDEIDPPGKGR-ARRGEPVPAPADMTLSHRLTHGRVRCCHARNPACGRCPILRWC 219

Qy 197 ----RGKEAPGRYPADPRRRAR--EBRLVALVLLGR 226

Db 220 PSYAAGETDPERARALLAYELKPGREULLLELLRAAR 255

```

RESULT 6
US-10-156-761-12127
: Sequence 12127, Application US/10156761
: Publication NO. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156,761
: CURRENT FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 12127
: LENGTH: 310
: TYPE: PRT
: ORGANISM: Streptomyces avermitilis
: US-10-156-761-12127

```

Query Match	10.9%	Score 184.5	DB 15	Length 310
Best Local Similarly	25.8%	Pred. No. 3.5e-09		
Matches 70	Conservative 44	Mismatches 110	Indels 47	Gaps 10

QY	1	RENAARDLPMNGE----	KDQYRVLVSEVLLQOQ----	RVEQALPYRRRLFERPTLKALAA	65
Db	65	RELAEEYYPVPAHPLELDRENSFOLLITLVLSAQOTDTRVQOTPR----	ALPAKPTPTEDILAAA	122	
QY	66	SLEEVLRVWQAGAGYYR-RAEHLRLRLARSV----	ELPPSFAELRGLPGLCPYTAANVAS	119	
Db	122	NPEEVEEILTRPGEFFPAKTKSVIGSKALVEEFGGEVPGRLIEDLVKLPQVGRKTAFAVVLG	181		
QY	120	IAEGEVAAVDQGVNRVRLVSLRFPARESPKEKELFALAOGLLPBGVDPGVWQNALMEL----	G	176	
Db	182	NAEGRGRCITVDTHFQGLVLRWQMTDEKDBDKLEAVGVGLFPR----	SEMTLSHHVLFPHG	233	
QY	177	ATVCLERKRPBGACPLGAFRCRCKEAPGRYPAPRKRRAKEERLVALVTLGRGVHLERLEG	236		
Db	238	RRICHARKPACGACPIAPLCPRAYGGEOTDEPKAKTKLYE-----	KG-----	279	
QY	237	RFOGLXGVLEPPEELRGREAPGYVRSPRLG	267		
Db	280	---GFGGQRINLPQAVLD----	AGGIAPAPLG	304	

RESULT 7  
US-09-738-626-3828  
US-09-738-626-3828  
; Sequence 3828, Application US/09738622  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGUCHI, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIRO

```

: APPLICANT: YOKOI, HARUHIKO
: APPLICANT: TATEISHI, NAKO
: APPLICANT: SENOH, AKIHIRO
: APPLICANT: IKEDA, MASATO
: APPLICANT: OZAKI, AKIO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-125
: CURRENT APPLICATION NUMBER: US/09/738, 626
: CURRENT FILING DATE: 2000-12-18
: PRIOR APPLICATION NUMBER: JP 99/377484
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: JP 00/159162
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 7059
: SOFTWARE: PatentIn ver. 3.0
: SEQ ID NO 3828
: LENGTH: 260
: TYPE: PRT
: ORGANISM: Corynebacterium glutamicum
US-09-738-626-3828
```

Query Match 8.7%; Score 147.5; DB 10; Length 260;

Best Local Similarity 29.0%; Pred. No. 8.6e-06;

Matches 56; Conservative 29; Mismatches 83; Indels 25; Gaps 8;

```

QY 26 DRYRVLYSEVLQO---TRVEQALPYRRFLERFPLKALAASLEEVLRWOGAGYR- 81
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 53 NPLELVATILSAOCTDVRVNOQVTP---ALFKRYPATDVAANDKRELEEFIRPTGYR 109
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 82 RAHLHRLARSV-----ELPPSFAELRGLPGIGPTAAVAASIAFGERYAAVDGNVRRY 136
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 110 KATSLGLGRLSLHDGQVPGTLELPGVGRKTANVGNAGVGCTIVDTFHGRLL 169
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 137 LSR--LFARSPKEKELFALAOGLPEGVDPGVW---NQALMELGATVCLPKRPGGACP 191
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170 VRRLKTLDEEDPYKVE-----KVMNELIEKPEMTFHSRLIRHGRRICSHRAACGACM 223
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 LGAFGR--GKEAP 202
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 224 LAADCPSPFGLGP 236
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 8

```

: Sequence 296, Application US/09912020
: Patent No. US20020045592A1
: GENERAL INFORMATION:
: APPLICANT: zyskind, Judith
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Trawick, John
: APPLICANT: Forsyth, R. Allyn
: APPLICANT: Froelich, Jamie M.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
: FILE REFERENCE: ELITRA.001DVI
: CURRENT APPLICATION NUMBER: US/09/912,020
: CURRENT FILING DATE: 2001-07-23
: PRIOR APPLICATION NUMBER: 09/492,709
: PRIOR FILING DATE: 2000-01-27
: PRIOR APPLICATION NUMBER: 60/117,405
: PRIOR FILING DATE: 1999-01-27
: NUMBER OF SEQ ID NOS: 485
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 296
: LENGTH: 211
: TYPE: PRT
: ORGANISM: E. Coli
US-09-912-020-296
```

Query Match 8.5%; Score 145; DB 9; Length 211;

Best Local Similarity 29.2%; Pred. No. 1.1e-05;

Matches 64; Conservative 33; Mismatches 88; Indels 34; Gaps 11;

```

QY 2 EAMRKALLMAYRE-NRPLPMRGEXDPYRVVLYSEVLQO---TRVEQALPYRRFLERFP 57
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3 KATRLRLTRKRNHNHPPTTELNFSSPELRLAVLSAQTAVSNKAT-----AKLYP 56
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 ---TKALAASLEEVLRWOGAGYR-RAHLHRLARSV-----ELPPSFAELRGLPG 108
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 57 VANTPAAMELEGEVGYKTYIKTIGLYNSKAENIMIKTIRLLQHNGEVPEDEAALEALPG 116
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 109 LGPYTAAVAASIAFGERYAAVDGNVRRVSLR-FAR---ESPKEKELFALAOGLPE-- 161
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 117 VGRKTANVVLNFAFGPTTAVDTHTFRCVCRTOFAPGRKNVEQVEKLL-----KVPAPF 171
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 GVDPGVWNOALMELGATVCLPKRPGACPLGAFGRKE 200
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 172 KVDCHHW---LILHGRTCIARKPRGSCIIEDLCYEKE 207
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 9

```

: US-10-128-714-3128
: Sequence 3128, Application US/10128714
: Publication No. US20030119013A1
: GENERAL INFORMATION:
: APPLICANT: Jiang, Bo
: APPLICANT: Hu, Wengq
: APPLICANT: Tishkoff, Daniel
: APPLICANT: Zamudio, Carlos
: APPLICANT: Eroschkin, Alexey M
: APPLICANT: Lemieux, Sebastien M
: TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
: FILE REFERENCE: 10182-018-999
: CURRENT APPLICATION NUMBER: US/10/128,714
: CURRENT FILING DATE: 2002-04-23
: PRIOR APPLICATION NUMBER: US 60/285,697
: PRIOR FILING DATE: 2001-04-23
: PRIOR APPLICATION NUMBER: US 60/287,066
: PRIOR FILING DATE: 2001-04-27
: PRIOR APPLICATION NUMBER: US 60/295,890
: PRIOR FILING DATE: 2001-06-05
: PRIOR APPLICATION NUMBER: US 60/303,899
: PRIOR FILING DATE: 2001-07-09
: PRIOR APPLICATION NUMBER: US 60/316,362
: PRIOR FILING DATE: 2001-08-31
: NUMBER OF SEQ ID NOS: 8603
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 3128
: LENGTH: 281
: TYPE: PRT
: ORGANISM: Aspergillus fumigatus
US-10-128-714-3128
```

Query Match 7.5%; Score 128; DB 15; Length 281;

Best Local Similarity 28.6%; Pred. No. 0.00066;

Matches 44; Conservative 26; Mismatches 66; Indels 18; Gaps 7;

```

QY 59 LKALAASLEEVLRWOGAGYRRAHLHRLARSV-----ELPPSFAELRGLPGIGPY 112
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 97 LENTLAWSPEKLNLITVGFHNHNTKYIAAAEILIDQYNSDIPSTAEELMKLPGVGP 156
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 113 TAAVAASIAFG-ERVAAVDGNVRRVLSRLPARSPKEKELFALA-QGILLPEGVDPGVW-- 168
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 157 MAYICSAANGKDGITVDVHVRT-TNLMGMHKTPTPESTRNALBSWLPFR-----DKWHE 211
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 -NQALMELGATVCLPKRPGACPLGA--FCRGR 199
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 212 INKLVLGQTVCLPVGRCGECDDLAGTKLCKSE 245
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

RESULT 10
US-10-128-714-8128
; Sequence 8128, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengli
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroszhkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8128
; LENGTH: 461
; TYPE: PRF
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8128

```

```

Query Match
Best Local Similarity 7.5%; Score 128; DB 15; Length 461;
Matches 44; Conservative 26; Mismatches 66; Indels 18; Gaps 7;
QY 59 KALAAASLEEVYRVNQGAGYRRRAHLRLARSV-----EELPPSFALRLPLGIPY 112
DB 277 LEETILAVSEPKNELRLRTGFFHNKTKYTKAAAEILRDYNSDIPSTAEELMKLPVGPK 336
QY 113 TAAAVASIAFG-ERVAVDGNVRVLSRLFARESPEKELEFALA-OGILPEGYDPGW-- 168
DB 337 MAYLCASAMGKDEGCGVHVHNRIT-TNLMGMHKTTPRETTRALASMLPR----DKMHE 391
QY 169 -NOALMELGATVCLPKRRPGACPLGA--FCRCGK 199
DB 392 INKLIVGLGQTVCLLPYGRRCGECDLAGTKLKSE 425

```

```

RESULT 11
US-09-840-743-11
; Sequence 11, Application US/09840743
; Publication No. US20030135890A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Choi, Yeonhee
; APPLICANT: Hannon, Mike
; APPLICANT: Okamura, Jack Kishiro
; APPLICANT: Tatarinova, Tatiana Valerievna
; TITLE OF INVENTION: The Regents of the University of California
; FILE REFERENCE: 023070-099910US
; CURRENT APPLICATION NUMBER: US/09/840,743
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 09/553,690
; PRIOR FILING DATE: 2000-04-21
; NUMBER OF SEQ ID NOS: 119
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 1332
; TYPE: PRF

```

```

; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: DMT3 (IDMT3)
US-09-840-743-11

```

```

Query Match
Best Local Similarity 6.6%; Score 112.5; DB 12; Length 1332;
Matches 59; Conservative 34; Mismatches 95; Indels 71; Gaps 12;
QY 86 LHLRLARS-----VEELPPSFAL-RLGLPGIGPTAAVASIACGERVAADVGR 135
DB 849 LNRVKKHGSIDLEWLEDPDPDKAKEXYLSINGLKSVCVRLSLHQIAFPVDVTVGR 908
QY 136 VLRLRF--ARESPEKELEFALAOGILPEGYDPGW-----NOALMEL-----GAT 178
DB 909 IAVRLGWVPLQPLPDELQMLLELYPLESVQXILWPRCKLDOKITIELHYHMTTGKV 968
QY 179 VCLPKRPGACPLGAFCR--GKEAPGRYPAP-----RRRAKERLYV---- 219
DB 969 FCTKVKRNCNACPMKACRHYSSARASRLALPEPESDRTSVMIHERSKRPVYVNR 1028
QY 220 -ALVLLGRKGVHLELEGRFOGLYCVLPPEELPGREAAFGVRSRPLGEVRIAL-THRR 277
DB 1029 PSFLYQEKQEAQRSQ-NCEPIIEEPASPEPEY-----IEHDIEDYPR 1071
QY 278 LRVFVRGALMEGEGEDPWK 296
DB 1072 DKNNV-----GTSEDPWE 1084

```

```

RESULT 12
US-09-840-743-2
; Sequence 2, Application US/09840743
; Publication No. US20030135890A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Choi, Yeonhee
; APPLICANT: Hannon, Mike
; APPLICANT: Okamura, Jack Kishiro
; APPLICANT: Tatarinova, Tatiana Valerievna
; TITLE OF INVENTION: The Regents of the University of California
; FILE REFERENCE: 023070-099910US
; CURRENT APPLICATION NUMBER: US/09/840,743
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 09/553,690
; PRIOR FILING DATE: 2000-04-21
; NUMBER OF SEQ ID NOS: 119
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1729
; TYPE: PRF
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: DEMETER (DMT)
US-09-840-743-2

```

```

Query Match
Best Local Similarity 6.5%; Score 111; DB 12; Length 1729;
Matches 51; Conservative 27; Mismatches 91; Indels 60; Gaps 9;
QY 60 KALAAASLEEVYRVNQGAGY-----YRRAHLRLARS-----VEELPPSFAL-RL 104
DB 1219 EAIRRASISISIAIRIGNNNNLAVRIKDFLEIVADHGIDLEWLRSPDKAKOYLL 1278
QY 105 GLDGLGPTTAAVASIAFGERVAVDGNVRVLSRLF--ARESPEKELEFALAOGILPE 161
DB 1279 SINGLIGKSVCVRLTLNHLAPVDVTVGRVIRAVRGWVPLQPLPSLQHLLELYPLE 1338
QY 162 GVDPGWV-----NOALMELGATVCLPKRRPGACPLGAFCR--GKEAPGR 204
DB 1339 SIQFLMPRLCKLDORTLYELAHYQLITFGKVFCTKSRPNCNACPMRGECRHHFASAVASAR 1398

```

QY 205 YPAERKRRAKEERLVALLLGKRGVHLERLEGRFOGLYGVLPPEELP 253  
Db 1399 LALP-----ADERSITSAT-----IPV-PPESFP 1422

## RESULT 13

US-09-840-743-8  
; Sequence 8, Application US/09840743  
; Publication No. US20030135890A1  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Choi, Yeonhee  
; APPLICANT: Hannou, Mike  
; APPLICANT: Okamoto, Jack Kishiro  
; APPLICANT: Tatarinova, Tatiana Valerievna  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Nucleic Acids That Control Plant Development  
; FILE REFERENCE: 023070-09991005  
; CURRENT APPLICATION NUMBER: US/09/840,743  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 09/553,690  
; PRIOR FILING DATE: 2000-04-21  
; NUMBER OF SEQ ID NOS: 119  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 1413  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: DMT2 (1DMT2)  
US-09-840-743-8

Query Match 6.5%; Score 110; DB 12; Length 1413;  
Best Local Similarity 23.3%; Pred. No. 0.29;  
Matches 42; Conservative 25; Mismatches 71; Indels 42; Gaps 6;

QY 60 KALAASLEEVLRVMQAGAY-YRREHLRLARSYE-----EL 96  
Db 886 KAIRADYKEVAETKSKRMHKLAEKLOYLTLNKKIMOGFLDLRVNDHGSIDLEMLDV 945  
QY 97 PPSFAE-LRGLPGIPYTAAVAASIAFGERYAAVDGNVRYLSRLP--ARESPKREL 151  
Db 946 PPDKAKEYLLSFNGGLSKSVCEVRLTLHLHAFPDVTNGRAVLAAGVPLQPLPESLQL 1005  
QY 152 FALAGGLPEGVDPGVW-----NQALMEL-----GATVCLPKRRCGACPLGACFR 197  
Db 1006 HLEWYPMLESIQKYLMPRLCKLDQKTLIELHYQMITEGKVFCTKSKPCNACPMKSGCR 1065

## RESULT 14

US-10-302-840A-6  
; Sequence 6, Application US/10302840A  
; Publication No. US20030134794A1  
; GENERAL INFORMATION:  
; APPLICANT: Madlison, Edwin L.  
; APPLICANT: Ong, Edgar O.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENCODING SEQUENCE OF A POLYPEPTIDE AND METHODS BASED THEREON  
; FILE REFERENCE: 24745-1622  
; CURRENT APPLICATION NUMBER: US/10/302,840A  
; CURRENT FILING DATE: 2003-01-24  
; PRIOR APPLICATION NUMBER: 60/332,015  
; PRIOR FILING DATE: 2001-11-20  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 635  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (104)...(332)  
; OTHER INFORMATION: CVSP17 protease domain

US-10-302-840A-6

Query Match 6.4%; Score 108.5; DB 12; Length 635;  
Best Local Similarity 27.6%; Pred. No. 0.14;  
Matches 66; Conservative 22; Mismatches 108; Indels 43; Gaps 10;

QY 105 GLPGL-SPYTAANAASINAFGERVAADVGNVRYLSRLFARESPEKELFALAGLPEGV 163  
Db 314 GEGKPGVYTRYAAVFKMDLOMSAASSREPSCRELLAMDPQELQADAAKCAFYARL 373  
QY 164 DPGVNNQALMELGATVCLPKRRCGACPLGAFCR--GKEAPGRYP-----APRRKRAK 214  
Db 374 CPGS-QGACARLAHQCCQRRRRRC-----GGFCSPGRCRGAEGRPQPLTAFTPVRSA 427  
QY 215 EERLVALVLGKGVHLERLEGRFOGLYGVLPPEELPGREAAFGVRSRPLGFRHALT 274  
Db 428 ELHSLAHTLGLILR-NAOELGPPRGILR--RLAPALP-----APALRESPLHPARERL 480  
QY 275 HR-----RLRVEVRGALWMEGEDPMKREPLRLMEKVLKALPLAHAGVPLP 323  
Db 481 HSGRNAQTRFPKRRPPEKRGANCCPGLEPLKQKLAALQ-----AHAMTILQVP 529

## RESULT 15

US-10-156-761-14029  
; Sequence 14029, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 14029  
; LENGTH: 549  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-14029

Query Match 6.1%; Score 104.5; DB 15; Length 549;  
Best Local Similarity 21.5%; Pred. No. 0.27;  
Matches 92; Conservative 46; Mismatches 134; Indels 155; Gaps 21;

QY 7 ALLAWYRENAAPLPMWGEKDPYRVIVSEV-----LLOQT-----RVEQAL----- 46  
Db 112 ALVETARTAGLPLVQLHREVPYTVTEVHTELVNGHYALLDRAEYVHNRCTEALLGGG 171  
QY 47 -PYRRFLERF---PLKLAASL-----EYLRVMQAGYTRAEHLRLARS 92  
Db 172 VPQVLRILADFSGNPFLFETADQRLLYAAGAPADTDPLQVWEGC-----RGQH----- 220  
QY 93 VEELPPSFAELRGLPGIPYTAAVAASIAF-----GERVAADV-----GNVRYL----- 137  
Db 221 -KDAPPAGTTIVDPGCGAGSVRAKVLVLRGVNPAVFVHIAAERAGSLAAYVMQAR 279  
QY 138 -----SRLPARESPEKELFALAOG---LDP-----EGVDP-GVW 168  
Db 280 OEERELAARGDFLTDLAGRIDADAPQAQAVLGFKPGAGPLPLVYVRLADGLSPGGGW 339  
QY 169 -----NQALMELGATVCLPKRRCGACPLGAFRCGEAPERYAPPKRRAKERLVAL 221  
Db 340 AVLARAVAEELASIGVPLLVGVPVEGRVPLLLGLRSSE-----ERSAVADVRAAA 390



```

QY      222 VLLGRKGVHLERLEGRFQGLYGVPLFPPEELPG-----REAFQVRS 263
      : | : : : | : | : |
Db      391 LRAQVERAGMQRPGAQ-----PPVYVGVAGGMAAASAGLRHAQTATAAGLSID 440
      : | : : : | : | : |
QY      264 RPLGEVVRHALTHRLRVEVVGALMEGEGEDP-----WKRP LPKLMEKVLKRALPL----- 313
      : | : : : | : | : |
Db      441 RPYWDAR-----RLDIDLTLWRLR-DDPDLAAFVDRRAIGPLRDHDNRSKPPLPTLQ 491
      : | : : : | : | : |
QY      314 --LAHAG 318
      : | : : : | : | : |
Db      492 TYLAHAG 498

```

Search completed: September 15, 2003, 14:02:48  
 Job time : 69 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 13:40:14 ; Search time 83 Seconds  
(without alignments)  
621.520 Million cell updates/sec

Title: US-09-938-901-2  
Perfect score: 1700  
Sequence: 1 VEAWRKALLAWYRENNARPLP.....VLKRLPLLAHGVPLDPA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq\_19Jun03:\*  
1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1700	100.0	325	24	ABP56413
2	492	28.9	387	23	ABBS4155
3	491	28.9	349	21	ABP79054
4	486.5	28.6	486	21	ABAB3881
5	484.5	28.5	535	18	AAW31912
6	484.5	28.5	535	19	AAW60243
7	478	28.1	365	23	ABBA9175
8	477	28.1	375	20	AAV35003
9	471.5	27.7	350	22	AAV69755

10	465.5	27.4	391	20	AAV09118
11	465	27.4	384	23	ABP27945
12	464.5	27.3	350	19	AAW60244
13	461	27.1	374	20	ABP30125
14	460.5	27.1	341	20	AAV37104
15	459.5	27.0	391	24	ABU01657
16	443	26.1	360	23	ABP39839
17	442.5	26.0	328	23	ABP65978
18	442	26.0	319	20	AAV09119
19	442	26.0	447	20	AAV86014
20	437	25.7	322	18	AAW27942
21	437	25.7	322	18	AAW25524
22	368	21.6	333	23	ABP27944
23	365	21.5	293	22	AAW25524
24	332.5	19.6	292	22	AAW41327
25	321.5	18.9	292	21	AAW41327
26	321.5	18.9	292	21	AAW41327
27	274.5	16.1	221	19	AAW11703
28	274.5	16.1	221	19	AAW11703
29	274.5	16.1	221	19	AAW11703
30	213.5	12.6	220	23	ABU51793
31	200	11.8	223	22	AAW86207
32	199	11.7	268	23	AAE15907
33	183.5	10.8	218	23	ABP25507
34	178.5	10.5	219	23	ABP48015
35	176.5	10.4	90	23	ABP35548
36	172.5	10.1	228	23	ABP65491
37	164.5	9.7	198	23	ABU51400
38	164.5	9.7	218	23	ABBS4387
39	164.5	9.7	304	18	AAW23155
40	164	9.6	210	22	AAW81939
41	164	9.6	224	22	AAW82795
42	162	9.5	210	23	ABP25506
43	161.5	9.5	143	22	AAU29544
44	160	9.4	224	23	ABP38380
45	159.5	9.4	367	21	AAW17150

## ALIGNMENTS

RESULT 1  
ABP56413  
ID ABP56413 standard; Protein: 325 AA.  
XX  
AC ABP56413:  
XX  
DT 12-MAR-2003 (first entry)  
XX  
DE Thermus thermophilus DNA repair enzyme MutY protein SEQ ID NO.2.  
XX  
KW Thermus thermophilus; DNA repair enzyme; enzyme; MutY; RecJ; RecF;  
KW TPCF; biochemistry; molecular biology; research.  
XX  
OS Thermus thermophilus.  
XX  
PN JP2002247985-A.  
XX  
PD 03-SEP-2002.  
XX  
PF 23-FEB-2001; 2001JP-0047762.  
XX  
PR 23-FEB-2001; 2001JP-0047762.  
XX  
PA (RIKA) RIKAGAKU KENKYUSHO.  
XX  
DR WPI; 2003-078924/08.  
XX  
DR N-PSDB; AB222143.  
XX  
PT A DNA repair enzyme gene, a protein, a recombinant vector, a  
PT transformant, preparation of DNA repair enzyme, repairing the error  
PT sequence of a DNA, and prevention of error synthesis of a DNA sequence  
PT

XX Claim 1; Page 14-15; 41pp; Japanese.  
PS  
CC AB222143 to AB22146 encode the *Thermus thermophilus* DNA repair enzymes  
CC MutY, RecF, and TRCF and TRCF from ABP56413 to ABP5416. The enzymes  
CC can be used as research reagents for biochemistry and molecular biology.  
XX  
SQ Sequence 325 AA;

Query Match 100.0%; Score 1700; DB 24; Length 325;  
Best Local Similarity 100.0%; Pred. No. 1.5e-153;  
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEAMRKALLAMYRENAAPLPWGEKDPYRVLYSEVLLQOTRVQDALPYRRFLERPTLK 60  
DB 1 VEAMRKALLAMYRENAAPLPWGEKDPYRVLYSEVLLQOTRVQDALPYRRFLERPTLK 60  
QY 61 ALAASLEEVLRWQAGYRRABEHLRLARSVEELPPSFAELRGLPGLPYTAAVAST 120  
DB 61 ALAASLEEVLRWQAGYRRABEHLRLARSVEELPPSFAELRGLPGLPYTAAVAST 120  
QY 121 AFGEERVAADVGNVRRVLSRLFARESPEKELFALAAGLLPEGVDPGVMQALMELGATVC 180  
DB 121 AFGEERVAADVGNVRRVLSRLFARESPEKELFALAAGLLPEGVDPGVMQALMELGATVC 180  
QY 181 LPRRPGACPLGAFGRKEAPGRYPAPRRRAKEERLVALVLLGRKGVHLERLEGRCFG 240  
DB 181 LPRRPGACPLGAFGRKEAPGRYPAPRRRAKEERLVALVLLGRKGVHLERLEGRCFG 240  
QY 241 LVGVPLPEPELPGRAAAGVRSRPLGEVRHALTHRLRVEYRGALMBEGEDPMKRP 300  
DB 241 LVGVPLPEPELPGRAAAGVRSRPLGEVRHALTHRLRVEYRGALMBEGEDPMKRP 300  
QY 301 KLMEXVLRKALPLLAHAGVPLPDA 325  
DB 301 KLMEXVLRKALPLLAHAGVPLPDA 325

## RESULT 2

ABBS4155  
ID ABBS4155 standard; Protein: 387 AA.

XX AC ABB54155;

DT 16-MAY-2002 (first entry)

XX DE Lactococcus lactis protein muty.

XX KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

XX OS Lactococcus lactis IL1403.

XX PN FR2807446-A1.

XX PD 12-OCT-2001.

XX PF 11-APR-2000; 2000FR-0004630.

XX PR 11-APR-2000; 2000FR-0004630.

XX PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.

XX PI Bojotline A, Sorokline A, Renault P, Ehrlich SD;

XX DR WPI; 2002-043418/06.

XX PT New nucleotide sequence useful in the identification or *Lactococcus*  
XX PT Lactis and related species -

XX PS Claim 6; SEQ ID No 857; 2504pp; French.

XX CC The present invention is related to a *Lactococcus lactis* nucleotide  
XX CC sequence (AB90521) and related proteins (ABBS3300-ABBS5621). The

CC nucleic acid sequence is useful in the detection and/or amplification of  
CC nucleic acid sequence, particularly to identify *Lactococcus lactis* or  
CC related species. The proteins of the invention are useful for the  
CC biosynthesis or biodegradation of a composition of interest. The  
CC invention helps research in lactic bacteria, particularly useful in the  
CC production of yogurt and cheese.  
CC Note: The sequence data for this patent is based on equivalent patent  
CC WO200177334 (published 18-OCT-2001) which is available in electronic  
CC format directly from WIPO at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX  
SQ Sequence 387 AA;

Query Match 28.9%; Score 492; DB 23; Length 387;  
Best Local Similarity 39.3%; Pred. No. 3.1e-38;  
Matches 108; Conservative 57; Mismatches 84; Indels 26; Gaps 10;

QY 1 VEAMRKALLAMYRENAAPLPWGEKDPYRVLYSEVLLQOTRVQDALPYRRFLERPTLK 60  
DB 8 IREFQDLSMTDDOKKPLPMKRTTEPKYKISTMSQOTVETVMPYERPMKYPTE 67  
QY 61 ALAASLEEVLRWQAGYRRABEHLRLARSVEE---LPPSFAELRGLPGLPYTAA 115  
DB 68 TLAQADDAELLLKMEGLGYSRARNLKTAAQEVVDKNGKFPDNLADILSLKIGPYTAA 127  
QY 116 AVASTAFGEERVAADVGNVRRVLSRLFARES---SPKKEELFALAAGLLPEGVD---PGVM 168  
DB 128 ALASTSFLAPPAIDGNLMRYTSRLFELOCDISKSSRKTF---DGYLRLLSKRRPGDF 184  
QY 169 NOALMELGATVCLPRPGACPLGAFRC---RCKEAPGRYPAPRRK-RAKEERLVALV 223  
DB 185 NOALMDLGLSTVCSKSPKCEACPLNYCAAAAGSKQL--NYPKTKIKKDKLYTAFAL 242  
QY 224 LGRKG-VHLERLEGRCFGILG-VPLPEPELPGRE 256  
DB 243 ENSIGEYTLERKPSK--GLADMTFPLTELPAAD 275

## RESULT 3

ABP79054  
ID ABP79054 standard; Protein: 349 AA.

XX AC ABP79054;

DT 07-MAR-2003 (first entry).

XX DE N. gonorrhoeae amino acid sequence SEQ ID 4638.

XX KW Antibacterial; infection; vaccine; gene therapy.

XX OS *Neisseria gonorrhoeae*.

XX PN WO200279243-A2.

XX PD 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-IB02069.

XX PR 12-FEB-2001; 2001GB-0003424.

XX PA (CHIR-) CHIRON SPA.

XX PI Fontana MR, Pizsa M, Maignani V, Monaci E;

XX DR WPI; 2003-058415/05.

XX PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a  
XX PT medicament for treating or preventing *N. gonorrhoeae* infection -

XX PS Disclosure; Page 528; 815pp; English.

XX CC The present invention relates to proteins from *Neisseria gonorrhoeae*.  
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention.

XX Sequence 349 AA;

Query Match 28.9%; Score 491; DB 24; Length 349;  
 Best Local Similarity 38.1%; Pred. No. 3,4e-38;  
 Matches 128; Conservative 56; Mismatches 104; Indels 48; Gaps 15;

QY 8 LLAAYENAR-PLPMGEKDPYRVLYSEVLTQOTRVQALPYRRLEPPTLKALAAS 66  
 Db LIMQOHGHNHPQ-VKNPYCVWLSEIMLQOTVAAVLDYPRLEKPTVQTLAAR 72  
 QY 67 LEEVLNMOAGYVYRRAEHLRLARSV---ELPPSFAELRGLPGIPYTAASVISA 121  
 Db 73 ODEVLSIMAGLYGYGRARNLHKAQOIVQGFSPSERKDLFTLCGVSTAAISAF 132  
 QY 122 FGRVAAVQGNVRRVLSRLFARE-SPEKE---LFLAAGLLP-EGVDPGVNQALMEL 175  
 Db 133 FNRRTLDGNVRRVLCRVFAQDGNPDKKFENSLWTLASLMPSENAIDPTTYOGLMDL 192  
 QY 176 GATVCLPRKPRGACPLGAFRCGKEAGRYPARKRRAKEER---LVALVLGRKG-VHL 231  
 Db 193 GATVCKRTKPLCQCPMADCEAKKONRTAELPRKTALEVOVLPLVLRNDDGAILL 252  
 QY 232 ER--LEGROGLGVPLPPEELPGRE---AAGVSRPLGE---VNAHTHRLRLRYEVR 283  
 Db 253 EKRTAGIMGLCYVPCF--ESLNGLSDFRAKLITMADMDDEQTALTHTRLHLMIT-- 308  
 QY 284 GALMEGEGEDP-----WKRP-----LPKLME 304  
 Db 309 ---PREGQMPSEHNSDGIWIKPGLKDYGLPKPLE 340

RESULT 4  
 AAB43881  
 ID AAB43881 standard; Protein: 486 AA.

AC AAB43881;  
 XX  
 DT 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:1326.

XX Human: cancer associated gene; cancer antigen; detection; cancer;  
 KW diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;  
 KW antidiabetic; antitubercular; antineoplastic; antiarthritic; antiviral;  
 KW antiinflammatory; antihypertensive; antiallergic; antibacterial; cardiant;  
 KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;  
 KW vasotropic; antiproliferative; antineoplastic; gene therapy; inflammation;  
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 KW allergic reaction; graft versus host disease; organ rejection;  
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KW neurological disease; drug screening.

OS Homo sapiens.

PN MO20055350-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000MO-US05882.

PR 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Ruben SM;

XX

DR WPI: 2000-587533/55.

DR N-PSDB; AAC78090.

PT Novel isolated nucleic acids comprising sequences encoding peptides

XX useful for treating or diagnosing e.g. cancer -

XX Claim 11; Page 1976-1978; 2352pp; English.

CC AAC7607 to AAC78448 encode the human cancer associated proteins given  
 CC in AAB43398 to AAB44238. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vulnerability; immunomodulator;  
 CC antidiabetic; antitubercular; antineoplastic; antiarthritic;  
 CC antiinflammatory; antihypertensive; antiallergic; antibacterial; antiviral;  
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
 CC nootropic; vasotropic; antiproliferative; antineoplastic; gene therapy;  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells; to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention.

XX Sequence 486 AA;

Query Match 28.6%; Score 486.5; DB 21; Length 486;  
 Best Local Similarity 36.6%; Pred. No. 1.4e-37;  
 Matches 136; Conservative 46; Mismatches 107; Indels 83; Gaps 13;

QY 1 VEAMRKALLMAYRENAARPLPMRG-----EKDPYRVLYSEVLTQOTRVQALPYRRF 52  
 Db 104 VTAFRGSLLSNYDEKRDLERRRAEDMDLRAAYVWVSEVWLQOTVATVNTYTGW 163  
 QY 53 LEEFPTLKALAASLEEVNLMAGLYGYSRGRRLQEGARKVYBELGSHMPTAETLQQL 106  
 Db 164 MQKWPRLQDLASASLEEVNLMAGLYGYSRGRRLQEGARKVYBELGSHMPTAETLQQL 223  
 QY 107 PGLGPTTAANAASIASEBERAAVADGNVRRVLSRLFA-----RESPEKELEFALAQGLPE 161  
 Db 224 PGVGRYTAGAIASTAFQATGVVDGNVRRVLCRVRAIGADPSSTLVSQLMGLAQOLV-D 282  
 QY 162 GVDPGVNQALMELGATVCLPKRPRGACPLGAFRCGKE----- 200  
 Db 283 PARPGDFNQAMELGATVCTPQRLCQCVESICRAKORVQEDQLASGLSPVVE 342  
 QY 201 -AP-----GRYPARRRRRAK---EERLVALVL-----LGRKGVHLE 232  
 Db 343 CAPRTGCGHCLPSPSEMDQTLGVNPPRRASRRPREESATCVLEQPALGQILVQ 402  
 QY 233 R-LEGROGLGVPLPPEELPGREDAF-----GVNSRPLGVRRAALTHRLRL 278  
 Db 403 RPNSSGLLAGLMEPSPVMEPSEQLORRALLQELQORXAGPLPATXRLHLEGEVHTFSIKL 462  
 QY 279 RVEYRGALMEGE 290  
 Db 463 TYQVYGLALEGQ 474

RESULT 5  
 AAM31912  
 ID AAM31912 standard; Protein: 535 AA.

AC AAM31912;

DT 27-MAR-1998 (first entry)

XX

DE Human mismatch repair protein Muty.  
XX Muty; hMYH gene; mismatch repair; non-polyposis colon cancer;  
KW xeroderma pigmentosum; gene therapy; diagnosis; human.  
XX  
OS Homo sapiens.  
XX MO9733903-A1.  
XX  
XX 18-SEP-1997.  
XX  
XX 11-MAR-1996; 96WO-US03239.  
XX  
XX 11-MAR-1996; 96WO-US03239.  
XX  
XX 11-MAR-1996; 96WO-US03239.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
XX Wet Y;  
XX  
XX WPI; 1997-470811/43.  
XX DR N-PSDB; AAT89194.  
XX  
XX  
XX DNA encoding human Muty protein - useful for detecting and treating  
PT mismatches in DNA especially in non-polyposis colon cancer and  
PT xeroderma pigmentosum  
PS  
PS Claim 1; Page 39-40; 59pp; English.  
XX  
XX  
XX This protein comprises human Muty, which is homologous to the  
CC Escherichia coli Muty protein involved in the pathway that corrects  
CC A/G and A/C mismatches as well as adenines paired with  
CC 7,8-dihydro-8-oxo-deoxyuracine in mutated DNA. Its amino acid  
CC sequence was deduced from a human cerebellum cDNA clone (see  
CC AAT89194). Muty polypeptides can be expressed in transformed  
CC host cells. The polypeptides, or the polynucleotides encoding  
CC them, can be used e.g. to repair oxidative damage to DNA, to  
CC prevent mutations from oxidative lesions, to treat genetic diseases  
CC related to a mutated hMYH gene, e.g. xeroderma pigmentosum and  
CC neoplasia, and to diagnose an abnormal transformation or a  
CC susceptibility to abnormal transformation of cells, particularly  
CC in a non-polyposis colon cancer.  
XX  
XX  
SQ Sequence 535 AA:  
  
Query Match 28.5%; Score 484.5; DB 18; Length 535;  
Best Local Similarity 36.6%; Pred. No. 2.5e-37;  
Matches 136; Conservative 46; Mismatches 107; Indels 83; Gaps 13;  
  
QY 1 VEAMRKALLAWYRENAAPLPWRG-----EKDPYRVLYSEVLLQOTRVBOALPYRRF 52  
DB 79 VTAFFRGSLLSWYDQERDLPMRRRAEDMDLDRAYAVWSEVMIQOTVATVINYTGW 138  
QY 53 LERFPTLKALAAASLEEVLRWOGAGYRRAEHLRLARS-VEEL---PPSFABL--RGL 106  
DB 139 MOKMPTLDDLASASLEEVNQIMAGLGYSRGRRLDEGARKVVEELGHNMPRTAETLQQL 198  
QY 107 PGLGPTAAVAASIAFGEVAAVDGNVRVLSRLFA-----RESPKEKEFLAQLGRLPE 161  
DB 199 PEGVGTAGALASIAFGATGVVDGNVAVLCRVRAIGADPSSTLVSOQLMGLAOQLV-D 257  
QY 162 GVDPGVWNOALMELGATVCLPKRRPGACPLGAFGRKE----- 200  
DB 258 PARPGDFNOAMELGATVCTPQRPLCSQCPVESLCAARQRVBOQLASGLSGSPDVEE 317  
QY 201 -APG-----GRYPAPKRRRAK---ERYALVLY-----IGRKGVHLERL 234  
DB 318 CAPNTGQCHLCLPSPBPMDQTLGVVNFPRKASRKPRRESSATVCELDQALGAOILLVQ 377  
QY 233 R-LEGRFOGLYGV--LFPPEELPGRBAF-----GVRSRPLGEVNHALTNRRL 278  
DB 378 RPNSGLLAGLWEPSPVWEPSEQLORKALLDELQRMAGRLPATNHLRHLEGVVHTFSHIKL 437

QY 279 RVEVRGALWEQE 290  
DB 438 TYQVYGLALEGQ 449  
  
RESULT 6  
ID AAM60243 standard; Protein; 535 AA.  
XX AAM60243:  
XX  
XX 19-AUG-1998 (first entry)  
XX  
XX  
XX Amino acid sequence of human MYH (hMYH).  
DE  
XX Human; MYH; diagnosis; cancer.  
XX  
XX Homo sapiens.  
XX  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 114 /note="nucleotides encoding this residue not given"  
FT  
XX JP10057076-A.  
XX  
XX 03-MAR-1998.  
XX  
XX 11-MAR-1997; 97JP-0099540.  
XX  
XX 11-MAR-1996; 96US-0013132.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX WPI; 1998-210407/19.  
XX DR N-PSDB; AAV35701.  
XX  
XX New isolated polypeptide - useful for, e.g. diagnosis of cancer  
PT  
PS Claim 1; Pages 20-21; 23pp; Japanese.  
XX  
XX The present sequence represents human MYH (hMYH). The MYH polypeptide  
CC can be used for diagnosing cancer, where the decreasing levels of the  
CC polypeptide are measured in a sample from a patient.  
XX  
XX  
SQ Sequence 535 AA:  
  
Query Match 28.5%; Score 484.5; DB 19; Length 535;  
Best Local Similarity 34.9%; Pred. No. 2.5e-37;  
Matches 130; Conservative 49; Mismatches 110; Indels 83; Gaps 11;  
  
QY 1 VEAMRKALLAWYRENAAPLPWRG-----EKDPYRVLYSEVLLQOTRVBOALPYRRF 52  
DB 79 VTAFFRGSLLSWYDQERDLPMRRRAEDMDLDRAYAVWSEVMIQOTVATVINYTGW 138  
QY 53 LERFPTLKALAAASLEEVLRWOGAGYRRAEHLRLARS-VEEL---PPSFABL--RGL 106  
DB 139 MOKMPTLDDLASASLEEVNQIMAGLGYSRGRRLDEGARKVVEELGHNMPRTAETLQQL 198  
QY 107 PGLGPTAAVAASIAFGEVAAVDGNVRVLSRLFA-----RESPKEKEFLAQLGRLPE 161  
DB 199 PEGVGTAGALASIAFGATGVVDGNVAVLCRVRAIGADPSSTLVSOQLMGLAOQLV-D 257  
QY 162 GVDPGVWNOALMELGATVCLPKRRPGACPLGAFGRKE----- 200  
DB 258 PARPGDFNOAMELGATVCTPQRPLCSQCPVESLCAARQRVBOQLASGLSGSPDVEE 317  
QY 201 -APG-----RYAPAPKRRRAK---ERYALVLY-----IGRKGVHLERL 234  
DB 318 CAPNTGQCHLCLPSPBPMDQTLGVVNFPRKASRKPRROSSATVCELDQALGAOILLVQ 377  
QY 235 ---EGRFOGLYGV--LFPPEELPGRBAF-----GVRSRPLGEVNHALTNRRL 278  
DB 378 RPNSGLLAGLWEPSPVWEPSEQLORKALLDELQRMAGRLPATNHLRHLEGVVHTFSHIKL 437







XX		WP1: 1999-256632/22.
DR	N-PSDB:	AAX34828.
XX		New adenine glycosylase from Streptococcus pneumoniae useful for
PT		diagnosing and treating diseases such as meningitis, pneumonia and
PT		endocarditis
XX		Claim 12; Page 8; 30pp; English.
PS		This represents a S. pneumoniae adenine glycosylase (muty) polypeptide.
CC		Muty can be used to vaccinate patients and raise an immune response
CC		against S. pneumoniae by administration of the protein. Muty protein may
CC		also be applied to implanted devices, wounds or skin to protect against
CC		or treat S. pneumoniae infections. Administration of muty polypeptide or
CC		the gene will also protect against H. pylori infection which causes
CC		diseases such as stomach cancer, ulcers and gastritis. Antagonists of
CC		muty may be administered to inhibit muty in an infected individual.
CC		Diseases such as meningitis, pneumonia, endocarditis, conjunctivitis and
CC		sinusitis may be diagnosed by detection of muty gene in an individual by
CC		RT-PCR, or detecting muty in a cell sample from a patient. The new muty
CC		protein is expressed by S. pneumoniae at specific stages of infection. It
CC		is important for bacterial viability as it contributes to the removal of
CC		oxidized guanidines from the genome, which can cause mismatches and to
CC		mutations. Muty and the sequences encoding it can therefore be used to
CC		diagnose or prevent bacterial infections without the use of antibiotics.
SQ	Sequence	391 AA;
Query Match	27.4%; Score 465.5; DB 20; Length 391;	
Best Local Similarity	34.2%; Pred. No. 1.le-35;	
Matches 124; Conservative	53; Mismatches 127; Indels 59; Gaps 11.	
OY	1	VEAMRKALLAWYRENAKPLPWGEKDPRVLYSEVLADGOTRYEQALPYRRFLEREPTLK 60
Db	17	VISIREKLAWYDEKKRDLPMKRKNPPIHINWSEIMLOQTQVDYIPIYERFLDFPIVE 76
OY	61	ALAAASLEEVLRVMOGAGYGRABEHLRLARSV-----ELDPPSFADLGLPGIGPYTA 115
Db	77	SLATAPBESLLKAMEGLDGYSRVRRMQAQAQQIMDFGQPDPNTYEGISLSKGIDPYTAG 136
OY	116	AVASTAEGERAAVDGNRRARLSLFANES-----PKKEELRALAQGLLPBGVPCGWMDA 171
Db	137	AISTIAELPPEPAVDGNYRWLARLFENVNHDIGIPSNKKIFQAMMELLINDRPEDFNQA 196
OY	172	LMEIGATVCLPKPRPGACSPGAFRCGE--APGRYP--APRKRAKKEERLVATLLARK 227
Db	197	LMDSGSIDESVYNRPRESPPVKDFSAAVQNGTMDRIYPIKSPKKKY-DITYKALVKNNSQ 255
OY	228	GVHL---ERLEGREFGLGYVP-----FPREE--LPGREAAFGVRSHPLG----- 268
Db	256	GQFLEKNESKLLAGFWHPFIEVDNQSQEGOFPLPHQVADESNSGPSRESFOQDYD 315
OY	269	-----VNHALTHRLRLVEV-----KQALWEGEGEDDPKRRPLPLYM 303
Db	316	LDVMDLVDCFTVOHVFSHRKMWOIVAGOVSDHFDSREVRWLSP-EEFKNYPLAKPO 374
OY	304	EKV 306
Db	375	OKI 377
RESULT 11		
ID	ABP27945	
ABP27945	standard; Protein: 384 AA.	
XX	ABP27945;	
AC		
XX	02-JUL-2002 (first entry)	
DT		
DE	Streptococcus polypeptide SEQ ID NO 5066.	
XX		
XX	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;	

Query Match	Best local similarity	Score	DB	Length	384
Matches 115	Conservative	67	Mismatches 125	Indels 60	Gaps 12
QY	1	VEAMRKALLAARYRNARPLPWRGKGEKRYRVLVSEVLQOQTFVEQALPYRRFLERFPTLK	60		
DB	17	IASRRRTLELVYDDEKRDLPWRRTNPNYIYVSEIMLQOQNVIVPIYKRFLEWPOIK	76		
QY	61	ALAAASLEEVLRVWQAGAGYTRARALHRLARASVE-----LPSEFAELRGILGCGPTAA	115		
DB	77	DLADAPPEOLIKAMEGIGYYSRVNNMOKAAQOVWVDEGIGFPHTYDDIASLKGIGPYTAG	136		
QY	116	AVASIAJGERVAAYADGVNRYRLSLRFARE-----SPKKEELFALAQGLLPGVDGVWNOA	171		
DB	137	AIAISITSEFLPPPAVDGVNRYMARLFEVYNDIGDPKNNKIFQAIMETLIDPDGRGENDQA	196		
QY	172	LMELGATVYCLPFRKRCGACPL-----GAFCRGKEAPGRYPAVR--KRAAKEERLVALVYLGR	226		
DB	197	LMDGJTDIESKAKTRPDPESPIRFNNAAYLNG--LYGKYYPIKPNKKRKKPRRIQAFVIRNQ	254		
QY	227	KGVHL--ERLEGR--FQGLYGVPLFPPEELPGRKAAV--GVNSRPL-----	266		
DB	255	NGQVILKKNKRGRLGFGFWSPFIIETETSPLSQDLDFDNDNSNPIMOTONETEEREYOLK	314		



```

Db      67 DLADAEEDLLKAMEGLGYSSRVNRNMQAAQVWVDGCGIFPHTYDDIASLKGIGPYTAG 126
QY      116 AVASINAFGERVAADVGNRVLSRLPARE-----SPKEKEIFALAOGLLPBGVDGWNQA 171
Db      127 ALASISFNNLPPEPVDNVMARLFEVNDIDDPKRRKLFQAIMELTIDPRGDPNQA 186
QY      172 LMEIGATVCLLPKPRGACPL-----GAFRCGRKEAPGRYPAPR-KRAKEERLVALVLTGR 226
Db      187 LMDLGDIDIESAKTIPRDESEPIREFNAYLNGTYS--KYP1KNTKKKPKR10QAFVLRNQ 244
QY      227 KGVHL--ERLEGR-FQGLYGVPLFPEBELPGREAPF-GVRSRPL----- 266
Db      245 NGQYLLEKNTKGRLLGFMFSFPIETSPLSQQLDFDDNQNPIIMQTONETFOREYQTK 304
QY      267 -----GEVRALTHRLRVE-VGAL-----W---EGEGEDPMKRPPLK 302
Db      305 PONTNHFPIKHTFESHOKWTIELIGVVKATDLPNAPHLKVAIEDFSLYPFPATPQKM 364
QY      303 MEKVLKR 309
Db      365 LETYLNQ 371

```

RESULT 14  
AAV37104 standard; Protein: 341 AA.

AAV37104;  
07-OCT-1999 (first entry)

Protein involved in intermediate metabolism of nucleic acids.

Vaccine: eye disease; conventional trachoma; nonendemic trachoma;  
paratrachoma; inclusion conjunctivitis; genital disease; peritphalitis;  
nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
batholinitis; pneumopathy; venereal lymphogranulomatosis.

Chlamydia trachomatis.

WO928475-A2.

10-JUN-1999.

27-NOV-1998: 98WO-IB01939.

04-NOV-1998: 98US-0107077.

28-NOV-1997: 97FR-0015041.

17-DEC-1997: 97FR-0016034.

(GEST ) GENSET.

Griffals R;

WPI; 1999-371125/31.

Genome sequence of Chlamydia trachomatis

Disclosure: Page 902: 1755pp: English.

AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
of Chlamydia trachomatis (see AA201425). The polypeptides can be used as  
vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
can also be used to control growth of the microorganism. Chlamydia  
trachomatis is responsible for a large number of diseases, e.g. eye  
diseases such as conventional trachoma, nonendemic trachoma,  
paratrachoma, and inclusion conjunctivitis; genital diseases such as  
nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
peritphalitis, batholinitis; pneumopathy in breast feeding infants;  
and venereal lymphogranulomatosis. The polypeptides of the invention  
may be of use in treating these diseases.

Sequence 341 AA;

Query Match 27.1%; Score 460.5; DB: 20; Length 341;  
Best Local Similarity 39.1%; Pred. No. 2,7e-35;  
Matches 111; Conservative 49; Mismatches 103; Indels 21; Gaps 9;

```

QY      6 KALLAWRENAARPLPMGKEDPYRVLYSEVLLQOTRYEQALPYRRFLERFPTLKALAA 65
Db      26 EALRSWPLESKRSPRMDSPTRVWVSEWMLQOTRAEVVYPTFLKMEFPTLQDLAAQ 85
QY      66 SLEEVLRWQAGAGYRRARHLRLARSVE-----ELPPSAELRGDLGEPYTAANAASI 120
Db      86 RESDVQLMGELGYSSARNRNLGAGARVITELFGEIYNDLALLSIGISYANALIAF 145
QY      121 AFGERVAADVGNRVLSRLPARESP-----KEKEIFALAOGLLPBGVDGWNQALMEL 175
Db      146 AFQOKNPADVGNVLRVMSRLFAIEESIDRNMTREITGLCESLTPD-QDPQVIAESFIEL 204
QY      176 GATVCLPFRPCGACPLGAFRCRG--KEAPGRYPAPRRKRAKEERL--VALVLGRKGVHL 231
Db      205 GARIC-KQPLCECPURSECTAYRGCTMEQYPV-RNTRRAISRLFRAVYIVLYKQDYLM 262
QY      232 ERLEGR--FQGLYGVPLF--PEBELPGREAPFGRSRPLGEVRH 271
Db      263 TKREKEIMAGLYEFPYQLPKEDCCDIKTIHLVQKDYGETLH 306

```

RESULT 15  
ABU01657 standard; Protein: 391 AA.

ABU01657;

11-FEB-2003 (first entry)

S. pneumoniae type 4 strain protein from coding region #1233.

Bacterial meningitis; pneumonia; sepsis; otitis media;

ear infection; antiinflammatory; antibacterial; immunostimulant;

auditory; respiratory; gene therapy; vaccine.

Streptococcus pneumoniae type 4 strain.

WO200277021-A2.

03-OCT-2002.

27-MAR-2002: 2002WO-IB02163.

27-MAR-2001: 2001GB-0007658.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Maignani V, Tettelin H, Fraser C;

WPI; 2003-040579/03.

N-PDB; ABX06945.

New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
or ear infection

Claim 1: SEQ ID No 2466; 56pp: English.

The invention relates to a protein comprising or having at least 50%  
identity to any of the 2469 amino acid sequences, identified in the  
specificiation (available on a computer readable format), or its fragment,  
expressed from 2469 of 2489 identified DNA coding regions from the  
Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
ABS56454. Also included are an antibody which binds one of the  
proteins, treating a patient by administering the protein, DNA or

CC antibody in a composition), a kit comprising first and second primers,  
CC which are the nucleic acid cited above or fragments between nucleotides  
CC 8-100 of a sequence not defined in the specification, for amplifying a  
CC target sequence contained within a Streptococcus nucleic acid sequence,  
CC where the first primer is substantially complementary to the target  
CC sequence and the second primer is substantially complementary to the  
CC complement of the target sequence, and where the parts of the primers  
CC having substantial complementarity define the termini of the target  
CC sequence to be amplified, assay comprising contacting a test compound  
CC with the protein, and determining whether the test compound binds to the  
CC protein and a Streptococcus pneumoniae bacterium, where one or more  
CC genes encoding the proteins has been rendered inactive. The proteins,  
CC nucleic acid molecules, antibody and compositions are useful as  
CC medicaments for treating or preventing a disease or infection due to  
CC streptococcus bacteria, particularly *S. pneumoniae*, such as pneumonia,  
CC sepsis, otitis media or ear infection. They are also useful in developing  
CC vaccines, diagnostics and antibiotics. The methods are useful for  
CC identifying immunodominant proteins. The present sequence is one of  
CC the 2469 proteins expressed by the identified coding regions from the  
CC genomic sequence.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from Wipo at  
CC ftp://wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 391 Aa:

CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

**SQ Sequence 391 AA;**

Query Match	27.08;	Score 459.5;	DB 24;	Length 391;

Best Local Similarity 33.98; Pred. NO. 4e-35;  
Matches 123; Conservative 53; Mismatches 128; Indels 59; Gaps 11.

Matches 123; Conservative 53; Mismatches 128; Indels 59; Gaps 117

[illegible]

Search completed: September 15, 2003, 13:50:45  
Job time : 86 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: September 15, 2003, 13:47:45 ; Search time 97 Seconds  
(without alignments)  
864.609 Million cell updates/sec

Title: US-09-938-901-2

Perfect score: 1700  
Sequence: 1 VEAMRKALLAWYRENAERPLP.....VLKRALPLLAHAGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_23:\*  
1: sp-archaea:\*  
2: sp-bacteria:\*  
3: sp-fungi:\*  
4: sp-human:\*  
5: sp-invertebrate:\*  
6: sp-mammal:\*  
7: sp-mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-vertebrate:\*  
13: sp-vertebrate:\*  
14: sp-unclassified:\*  
15: sp-virus:\*  
16: sp-bacteriap:\*  
17: sp-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	669.5	39.4	363	16	09RS42
2	552.5	32.5	358	16	08G240
3	551.5	32.4	375	16	08YFS8
4	531	31.2	366	16	092RF0
5	520	30.6	349	16	09AB55
6	515.5	30.3	367	16	08UH85
7	514.5	30.3	366	16	0985U7
8	505	29.7	368	16	08DUJ3
9	503	29.6	349	16	09JY77
10	501.5	29.5	369	16	031584
11	498	29.3	346	2	09ZEF41
12	498	29.3	349	16	09JYW9
13	497.5	29.3	372	16	09KEC2
14	492.5	29.0	354	16	08CV61
15	492	28.9	387	16	09CH97
16	491.5	28.9	350	16	08Z3U0

17	491	28.9	371	16	08ZHE8	08zne8 yersinia pe
18	491	28.9	415	16	08CZV0	08czv0 yersinia pe
19	484.5	28.5	362	16	08Y2D5	08y2d5 ralsstia s
20	484.5	28.5	521	4	09UBP2	09ubp2 homo sapien
21	484.5	28.5	522	4	09UF4	09uf4 homo sapien
22	484.5	28.5	532	4	09UF5	09uf5 homo sapien
23	484.5	28.5	535	4	015830	015830 homo sapien
24	484.5	28.5	536	4	09UIF6	09ulf6 homo sapien
25	484.5	28.5	546	6	09UIF7	09ulf7 homo sapien
26	483	28.4	365	16	09ZAX1	09zax1 listeria in
27	478.5	28.1	353	16	09KUR3	09kur3 vibrio chol
28	478	28.1	355	16	08Y6J5	08y6j5 listeria mo
29	477.5	28.1	355	16	09HU37	09hu37 pseudomonas
30	477	28.1	369	16	09Z8E1	09z8e1 chlamydia p
31	476	28.0	339	16	09PM55	09pm55 campylobact
32	472.5	27.8	365	16	08EBX7	08ebx7 shewanella
33	470.5	27.7	360	16	08FE21	08fe21 escherichia
34	469.5	27.6	350	16	08XCS8	08xcs8 escherichia
35	468.5	27.6	345	16	08DSD4	08dsd4 streptococc
36	467	27.5	345	16	099T11	099t11 staphylococ
37	466	27.4	374	16	08N2I9	08n2i9 streptococc
38	466	27.4	384	16	08K5Y9	08k5y9 streptococc
39	465	27.4	345	16	08NVU5	08nvu5 staphylococ
40	465	27.4	374	16	099Y77	099y77 streptococc
41	462	27.2	374	16	08E3J9	08e3j9 streptococc
42	461	27.1	374	16	08DXV0	08dxv0 streptococc
43	461	27.1	515	11	099P21	099p21 mus musculu
44	460.5	27.1	516	11	08R5G2	08r5g2 rattus norv
45	459.5	27.0	361	16	097Q13	097q13 streptococc

## ALIGNMENTS

RESULT 1  
Q9RS42 PRELIMINARY; PRT; 363 AA.  
ID Q9RS42  
AC Q9RS42;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE A/G-specific adenine glycosylase.  
GN DR2285.  
OS Deinococcus radiodurans.  
OC Bacteria: Deinococcus-Thermus; Deinococci; Deinococcales;  
OC Deinococcaceae; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RI;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hikey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.,  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans RI.";  
RL Science 286:1571-1577(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RI;  
RX MEDLINE=21475757; PubMed=11591657;  
RA Li X., Lu A.L.,  
RT "Molecular Cloning and Functional Analysis of the MutY Homolog of  
RT Deinococcus radiodurans";  
RL J. Bacteriol. 183:6151-6158(2001).  
DR EMBL: AE002060; AAF11831.1; -;  
DR EMBL: AF377342; AAL26976.1; -;  
DR HSP: P17802; IMUN.  
DR TIGR: DR2285; -.

DR InterPro: IPR004035; EndoIII\_FCL.  
DR InterPro: IPR003265; Endo\_3c.  
DR InterPro: IPR003651; Res\_bind.  
DR InterPro: IPR005760; MutY.  
DR InterPro: IPR001680; WD40.  
DR Pfam: PF00730; HhH-GPD; 1.  
DR SMART: SM00478; ENDO3C; 1.  
DR SMART: SM00525; PES; 1.  
DR TIGRfams: TIGR01084; muty; 1.  
DR PROSITE: PS00764; ENDOUNCLEASE\_III\_1; 1.  
DR PROSITE: PS00678; WD\_REPEATS\_1; 1.  
KW Complete proteome.  
SQ SEQUENCE 363 AA; 3938 MW; 1D45C6A09A2BE2F CRC64;

	Query Match	Similarity	Score	669.5	DB	Length	363
	Best Local	Similarity	46.2%	Pred. NO.	4.7e-42		
	Matches	154	Conservative	48	Mismatches	96	Indels
						35	Gaps
							11.
QY	1	VEAWRKLLAWYRENNARFLPMR----	GKEDKYRVLYSEVLLQOTRVEQALPYRRFLRFE	56			
Db	17	VGALRRDILLGFMFDAGNRGLPMRLDEGRNDYRWVVAWILLIQQVANGLGLEFFLEAF	76				
QY	57	PTLKALAASLEEVLRWOGAGYRRRAEHLRLARSVEE--LPSEAEIRGLPGIGPYTA	114				
Db	77	PTYQALAAAPDDVALKAMEGCGYYARAARNLRAAALIDEOGFPODYAGMLLPGVGYTA	136				
QY	115	AAVASTAFGERVAADVGNRRVLSRLFRRESPEKKEKELFALNQGLPEGVDAVGNQALME	174				
Db	137	AAVSSILALGERRAAYDGNVRRVLSLRLEAHPSDQWVEQADRLD-DPARRGAMNEAYMD	195				
QY	175	LGAIVVCLPKPRBCGACPLGAFCRGKE--APGRYPAPRRK-RAKEERLALVLLGKRGVHL	231				
Db	196	LGATITCYPKSACDRCPSAHCAAYQLQGPQDFPAPKAPQAREVRAVALLIGDEAVYL	255				
QY	232	ERLEGR--FQGLGYVPLPEPPEELPGREAAFGVRSR-----PLGEVRHALTHRLRL	279				
Db	256	EKREGSLTGLGFLGPL--EEIGARETAADALARLOARLGAEVKEKCTSTVOHGMYHRLIS	312				
QY	280	VEVKGALMEGEGEDPMKRP-----LPKLMKEYL	307				
Db	313	VEV-----YRAEADRP-RQVPVGCALSTRIDHNL	340				

RESULT 2			
ID	Q8G240	PRELIMINARY;	PRT; 358 AA.
AC	Q8G240;		
DT	01-MAR-2003 (TREMblrel. 23, Created)		
DT	01-MAR-2003 (TREMblrel. 23, Last sequence update)		
DT	01-MAR-2003 (TREMblrel. 23, Last annotation update)		
DE	A/G-specific adenine glycosylase.		
GN	MUTY OR BR0493.		
OS	Brucella suis.		
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;		
OC	Brucellaceae; Brucella.		
OX	NCBI_TaxID=29461;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=1330 / Biovar 1;		
RX	MEDLINE=22247741; PubMed=12271122;		
RA	Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,		
RA	Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,		
RA	Daugherthy S.C., Debroy R.T., Durkin A.S., Kolonay J.F., Madupu R.,		
RA	Nelson W.C., Ayojele B., Kraul M., Shetty J., Malek J., Van Aken S.E.,		
RA	Redmuller S., Mettelin H., Gill S.R., White O., Salzberg S.L.,		
RA	Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;		
RT	"The Brucella suis genome reveals fundamental similarities between		
RT	animal and plant pathogens and symbionts."		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).		
DR	EMBL; AF014358; AAN29436.1; -.		
DR	TIGR; BR0493; -.		
KW	Complete proteome.		
SO	SEQUENCE 358 AA; 39047 MW; E147F8C3CEDB74E7 CRC64;		

Query Match	32.5%;	Score 552.5;	DB 16;	Length 358;
Best Local Similarity	38.4%;	Pred. No. 2.4e-33;		
Matches 134;	Conservative 53;	Mismatches 111;	Indels 51;	Gaps 12;

Qy	8	LWAYREARPLPMR-----GE-KPYVULSEVLLOOTRYEOALPYRRERLEPPT	58
Dd	7	LLRRYDRIHNR/LPKRVTPVDAAKGDVADPRVWMLSEIMLOOTYEAVASYFLRFIERMPT	66
Qy	59	LKALAASLEEV/LRWOGAGYYRAAEHLRLARSY----BELPPSFELGDLGLGYT	113
Dd	67	VRAAKASEDDIL/KAMGLGYYSRARINLKKCADIVAAEHGEFSPSAAGLKELPEIDGYT	126
Qy	114	AAAASAFAEGRVAAYDGNVRVYSRLPARESP---KELEFLAALOGLLPGVDPGVNNQ	170
Dd	127	SAAIAAIIFGQVAVVDGNSVERVYSRLYAIDTPLPVAKAOICALLMGMTTPPD-RGGDPAQ	185
Qy	171	ALMEIGATVCL/PKRRCGCAP/LGAFCRG--KEAPGRYP--APRKRAKEERLVALVLLGR	226
Dd	186	AAMDLCGATICP/RPRACALCPLNMGICALCERDEDPVKAPKAEKPVRTGAAFIANGD	245
Qy	227	KGVHERLEGFGOGUYGPLFPPELPFG-----REAFFGVSRLPGLVEVRHAL	273
Dd	246	GSVYLIRKKG---BGL-----LAGMTPEVPSGMTARTIDSGATYNAAFPASAMNPISGITHYVF	299
Qy	274	THRLRAYEVRGA---LMEGEGEDPWKRP-----LPKLMEKLYLRKALP	312
Dd	300	THEFLRISVYRASNVNRKQANNEGWMSTEEELCGEALLPIVMKKAIALAAP	348

0Y	8	LLAWRENARDLPMW-----GE-KDPYRVLYSEVLLQOTRVQALPYRRRLERPEPT	58
----	---	---	----

```

Db      24 LIRRYDRHHHRLVPRVTPVDAKGDVADPYRVMLSEIMLQOTVEAVKSYLRFLEMP 83
Qy      59 LKALAAASLEEVLRVMQAGYRRAEHLHLARSV-----ELPPSPALGLGLGYT 113
Db      84 VRAAKASSEDDILAMAGLGYSSRAHNLKCCADIVVAEHGGEFPKSAAGLELPGIGYT 143
Qy      114 AAAVASIAFGERRVAAGNVRVLSRLFARES-KEKELFALAOGLLPGVDPGVNQ 170
Db      144 SAATAALAFGEQVAVVGNVERVSRLLAIDTLPVAKAQIRALMGQTPPD-RPGDPAQ 202
Qy      171 ALMEIGATVCLPKRRPGACPLGACRG--KEARGRP-APRRRAKEERLVALVLGR 226
Db      203 AMMDLGATICTPRRACALCPLNKGCIALCEPDEDPVAKPKAEKPYRTGAFIALAGD 262
Qy      227 KGVHLELRGFRGQGLVGPFPPEELPG-----REAFGVRSPLGEVRNAL 273
Db      263 GSVYLRKRKG--EGL---LAGMTEVPGSGWTARIDDAIVNAAPFSAAMTPSGTITV 316
Qy      274 THRRLRYEVGGA---LWEGEGEDPWKR- -LPRIMEKVKRLALP 312
Db      317 THEFLRLSYRASNVKQANEGWSTPEELGELPTVMKKAIAALP 365

```

## RESULT 4

```

Q92RF0 ID 092RF0 PRELIMINARY; PRT; 366 AA.
AC 092RF0:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Probable A/G-specific adenine glycosylase protein.
GN MMY OR R00928 OR SMC00452.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021.
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kise E., Lelaure V., Masuy D.,
RA Pohl T., Portetle D., Puehler A., Purnelle B., Ramspeger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591785; CAC45500.1; -.
DR InterPro: IPR004035; EndoIII_FCL.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003651; Fes_bind.
DR InterPro: IPR000445; Hnh.
DR InterPro: IPR005760; Muty.
DR Pfam: PF00633; HNH_1.
DR Pfam: PF00730; Hnh-GPD_1.
DR SMART: SM00478; ENDO3c_1.
DR SMART: SM00525; FES; 1.
DR TIGRFAMs: TIGR01084; muty; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
KM Complete proteome.
SQ SEQUENCE 366 AA; 39944 MW; 698224AA25089831 CRC64;

```

## Query Match

Best Local Similarity 31.2%; Score 531; DB 16; Length 366;  
Matches 136; Conservative 41; Mismatches 125; Indels 60; Gaps 11;

```

Qy      8 LLAMYRENARPLPWR-----GEKDPYRVLSVLLQOTRVEQALPYRRLEFP 58
Db      14 LLEMYDRHHHDLPMWRPPAARAKGAVADPYRVMLSEVLLQOTVQAVKAEKFLALMP 73
Qy      59 LKALAAASLEEVLRVMQAGYRRAEHLHLARSV-----ELPPSPALGLGLGYT 113

```

```

Db      74 VGDIAADTDEVMAMAGLGYRARARNLKKCAENARHDGRFPDSEEGALALPGIGYT 133
Qy      114 AAAVASIAFGERRVAAGNVRVLSRLFARES-KEKELFALAOGLLPGVDPGVNQ 170
Db      134 AAAIAALAFNPNASAVLDGNGVERVSRLLAIVETPLPAKPEMRALVQALTPAD-RPGDPAQ 192
Qy      171 ALMEIGATVCLPKRRPGACPLGACRGCKEAKGRNPAPRRRAKEERLVALVLGRGVN 230
Db      193 AMMDLGATICTPRRACSLCPFTDCLAKTADEPTFPKAKKEPL-----RUGAFVA 248
Qy      231 LERLEGRF-----QGLYGVPLFPPEELPGRE-----AAFGVRSPLGEVRH 271
Db      249 VDGLEAVYLRKRPEFTGLG-----GMTEVPGTDWTSRRDGTSIDAHPPEAMEPCGTYNH 304
Qy      272 ALTHRLRYEV-----RGALWEGEGE-DPWKR- -LPRIMEKVKRLALP 316
Db      305 VTFHEFLHSYFARVGRADIGEARDTSGMWEPLASIAQALPTVMKKAIAALP 364
Qy      317 AG 318
Db      365 AG 366

```

## RESULT 5

```

Q9AB55 ID 09AB55 PRELIMINARY; PRT; 349 AA.
AC 09AB55:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE A/G-specific adenine glycosylase.
GN CC0377.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=15892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RA MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin R.J., Gilm M.L., Hatt D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouiri H., Shetty J., Berry K.,
RA Ueberlack T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AB005710; AAK2364.1; -.
DR HSSP; P17802; IMON.
DR TIGR; CC0377; -.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003651; Fes_bind.
DR InterPro: IPR000445; Hnh.
DR InterPro: IPR005760; Muty.
DR Pfam: PF00633; HNH_1.
DR Pfam: PF00730; Hnh-GPD_1.
DR SMART: SM00478; ENDO3c_1.
DR SMART: SM00525; FES; 1.
DR TIGRFAMs: TIGR01084; muty; 1.
KM Complete proteome.
SQ SEQUENCE 349 AA; 37615 MW; DD96F16FB2A4ABC2 CRC64;

```

## Query Match

Best Local Similarity 30.6%; Score 520; DB 16; Length 349;  
Matches 141; Conservative 33; Mismatches 137; Indels 36; Gaps 10;

```

Qy      2 EAWKALLAMYRENARPLPWR-----GEKDPYRVLSVLLQOTRVEQALPYRRLEFP 52
Db      5 DALRSALLAWYDAQARDLAWRVGPAERAGVRSPPYRVMLSEVLLQOTVPHARPYPLSF 64

```

```

QY 53 LERPEFLKALAAASLEEVLRWOGAGYVRAEHLRLARSVE-----LPPSFAELRGIP 107
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 65 TORWPTVLDAAVEDGDLMAAAGLGYYARARNLLACARAAVANDGAVFPGTEGRLALP 124
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 GLGPTAAAVASIAIGERAAVNDGNVRYLSLRFRESF---KEKELFALAGLLPEGV 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 125 GVGATTAATAAVALIADRAANVVDGNVERVMSRLFAVEAPMPDSKPELALADLTVD-D-R 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 165 PGVWMOALMELGATVCLPKRPRCGACPLGAFCRG--KEAPGRYPAPRRKRAKEERL-VAL 221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 184 PGDAWQALMDLGATCTCKRKGRPLCDRCSPYSKCAAVGGAPEYTPKTKKADPRKHGVAY 243
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 222 VLLGRKGVHLERL--EGRFQGLGYVPL-----FPPEELPGREAAFGVRSRPLGEVRR 272
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 244 VLTGDEVALVRRPPEKGLLGMLGLPTSDWMTASTYDDAEAVA-AAPLAAAMRDLGAVEHV 302
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 273 LTHRLRLREYVRCALMEGEGEDPMK-----RPLPKLMEKYLKRALPL 314
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 303 FTHESLTLRVFAADGANDGDFVWTPREGIGALPVSFLKAAAMAQRLL 349

```

## RESULT 6

```

Q80H85 PRELIMINARY: PRT: 367 AA.
AC Q80H85;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE A/G-specific adenine glycosylase.
GN MUTY OR ATU0798 OR AGR_C1460.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;

```

```

RN [1]
RP MEDLINE-21608550; PubMed=11743193;
RX MEDLINE-21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitaajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eilen J.A., Karp P.D., Boyce D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li W.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Seemphilmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP MEDLINE-21608551; PubMed=11743194;
RX MEDLINE-21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quinello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houtille K., Gordon J., Vaubin M., Iatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Marzelz B.,
RA Flanagan C., Crowell C., Gursun J., Lomo C., Seer C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL: AE009046: AAL41814.1; -.
DR EMBL: AE008012: AAK86607.1; -.
DR InterPro: IPR004035; EndoIII_FCL.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003651; Fes_bind.
DR InterPro: IPR005760; MutY.
DR Pfam: PF00730; Hnh-GPD; 1.
DR SMART: SM00478; ENDO3C; 1.
DR SMART: SM00525; FES; 1.
DR TIGRFAMS: TIGR01084; muty; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.

```

```

KW Complete Proteome.
SQ SEQUENCE 367 AA: 40056 MW: 5237C00E3FEF994A5 CRC64;

```

```

Query Match 30.3%; Score 515.5; DB 16; Length 367;
Best Local Similarity 37.1%; Pred. No. 1.4e-30;
Matches 130; Conservative 55; Mismatches 124; Indels 41; Gaps 13;

```

```

QY 3 AMRKALLAMYRENAARLPW-----RGEK-DPRVLVSEYLLQOTRVEQALPYRRRL 53
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 9 SYAQMLLAWYDRHNRHELPRRTSPAMAAARGKADRPVNVMLSEYMLQOTTVQAVKRYELKFL 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 54 ENFPTLKALAAASLEEVLRWOGAGYVRAEHLRLARSVE-----LPPSFAELRGIP 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 69 AAMPVVDLDTAPVEDVMAAMAGLGYARARLKCACAAVAREHGGVPEPTDEEGUKQLPG 128
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 109 LGPTAAAVASIAIGERAAVNDGNVRYLSLRFRESF---KEKELFALAGLLPEGV 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 129 IGDYTSAAVAATAFNQAAVMDGNVERVYISRLFALDAPLPGRKPMKAKVALLTP-AERP 187
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 166 GYWMOALMELGATVCLPKRPRCGACPLGAFCRG--KEAPGRYPAPRRKRAKEERL-VALV 222
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 188 GDFQAQAMMDLGATICTPKRPACALCFPNAGCLALTRDEBERPPVAKAKKAPVRLGAAYF 247
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 LIGRKGVLHRL--EGRFQGLGYVPLP-PEELG-----RRAAGVSRPLGEVRRALT 274
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 248 AVNTGELLRLRRIDSGILGTMTEVPTAAMTARMGTEVSHAPAAAGMAGVIGHVFT 307
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 HRRLRVEY-RGALMEG---EGEDPMKRP-----RPLMKYLRKALP 312
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 308 HFELRLTYRAQVPPGLGTGPDGMEPEVTNIDQAQLPYMKKVIARALP 357

```

## RESULT 7

```

Q985U7 PRELIMINARY: PRT: 396 AA.
AC Q985U7;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
DE Adenine glycosylase.
GN ML17523.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;

```

```

RN [1]
RP MEDLINE-MAFR303093; PubMed=11214968;
RX MEDLINE-21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003012; BAB53965.1; -.
DR InterPro: IPR004035; EndoIII_FCL.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003651; Fes_bind.
DR InterPro: IPR005760; MutY.
DR Pfam: PF00730; Hnh-GPD; 1.
DR SMART: SM00478; ENDO3C; 1.
DR SMART: SM00525; FES; 1.
DR TIGRFAMS: TIGR01084; muty; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
KW Complete proteome.
SQ SEQUENCE 396 AA: 43649 MW: 2D3D11219667049E CRC64;

```

```

Query Match 30.3%; Score 514.5; DB 16; Length 396;
Best Local Similarity 37.6%; Pred. No. 1.9e-30;
Matches 130; Conservative 45; Mismatches 124; Indels 47; Gaps 10;

```





Db 253 EKRPAKIGWGLVCVPCF--ESLNGLSDFAKFSLTWADMDEQATLTHRLTLTLTPF 310  
 QY 284 GAWL--EGEGEDPPKRP-----LPKIME 304  
 Db 311 EAQWPSPSPDGIMIKRAHLKDGGLRPLE 340

RESULT 10  
 031584  
 ID 031584 PRELIMINARY; PRT: 369 AA.  
 AC 031584;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE YFHQ protein.  
 GN YFHQ.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=938437;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Biotin A., Borchert S.,  
 RA Borris R., Boursier L., Brans A., Brun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.U., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Deniot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Eutlin K.D., Errington J., Fabret C., Ferrari E., Fougere D.,  
 RA Fritz C., Fujita M., Fujita Y., Funo S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Hage K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holstappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetlelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schnoeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yaman K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 RT subtilis".  
 RL Nature 390:249-256(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97101647; PubMed=8946165;  
 RA Yamamoto H., Uchiyama S., Sekiguchi J.;  
 RT "Cloning and sequencing of a 27.8-kb nucleotide sequence of the 79  
 RT degrees-81 degrees region of the Bacillus subtilis genome containing  
 RT the sspE locus".  
 RL DNA Res. 3:257-262(1996).  
 DR EMBL: D89108; CAB12691.1; -;  
 DR EMBL: D85082; BAA24483.1; -;  
 DR HSSP: P17802; 1MOV.  
 DR InterPro: IPR003265; Endo\_3c.  
 DR InterPro: IPR003651; Fes\_bind.  
 DR InterPro: IPR000445; HHH.

DR InterPro: IPR005760; NucY.  
 DR Pfam: PF00633; HHH; 1.  
 DR Pfam: PF00730; HHH-GPD; 1.  
 DR SMART: SM00478; ENDO3c; 1.  
 DR SMART: SM00525; FES; 1.  
 DR TIGRFAMs: TIGR01084; nucY; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 369 AA; 41963 MW; FE1FE70018579B1 CRC64;

Query Match 29.5%; Score 501.5; DB 16; Length 369;  
 Best Local Similarity 36.1%; Pred. No. 1,6e-29;  
 Matches 112; Conservative 47; Mismatches 108; Indels 43; Gaps 6;

QY 1 VEAMRKALLAWRENAARPLPMWGEKDPYRVLSVTLQOTVEQALPYRRFLERPLPK 60  
 Db 13 IQQFDDLLSWFERORQVLPWREDDPYKVVWSEVMLQOTREVIPIYFLRFVQPFIVE 72  
 QY 61 ALAASLEBVLKVMQAGYVYRAEHLHRLARSVE-----LPPSFAELRGLPGLYTAA 115  
 Db 73 ALADADEKVLKAWEGGLGYYSVRNLQSAVKEVQOEYGVGPPEKDEGLKGVGYTTKG 132  
 QY 116 AVASTAFGEERVAADVGNRVLSRLFA---RESPEKEFLAQLPEGVDGVMNQ 171  
 Db 133 ALASTAYNKKPIPAVDGNVWRVMSRLTSIMDDIAKPKRTTFEDAIRAFISKEKSEFENOG 192  
 QY 172 IMELGATVCLPKRPKPCACPIGAFRCRKEAPRGVAPAPKRAKE---ERLVATVLLGRKG 228  
 Db 193 LMEIGALCTKRSPLCCLPVOQHCSAFEEGTRELPYKSKKKPGIKTMAIVLTDEDG 252  
 QY 229 ---VHLERLEGRFDGLGVPLPPEELPGRBAEGR-----SRP 265  
 Db 253 QVYHKRPSKLLANLW-----EPNLETKGIKTERPQLAFLENEXGIQADISDL 304  
 QY 266 LGEVHALTH 275  
 Db 305 QGVVEHVFTH 314

RESULT 11  
 092F41  
 ID 092F41 PRELIMINARY; PRT: 346 AA.  
 AC 092F41;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Adenine glycosylase.  
 GN NucY.  
 OS Neisseria meningitidis.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=487;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96118708; PubMed=7496539;  
 RA Kroll J.S., Langford P.R., Wilks K.E., Kell A.D.;  
 RT "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct  
 RT from the eukaryotic enzyme, and not so rare after all".  
 RL Microbiology 141:2271-2279(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98084476; PubMed=9423860;  
 RA Wilks K.E., Dunn K.L., Farrant J.L., Reddin K.M., Corringe A.R.,  
 RA Langford P.R., Kroll J.S.;  
 RT "Periplasmic superoxide dismutase in meningococcal pathogenicity".  
 RL Infect. Immun. 66:213-217(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98445380; PubMed=9770495;  
 RA Kroll J.S., Wilks K.E., Farrant J.L., Langford P.R.;  
 RT "Natural genetic exchange between Haemophilus and Neisseria:  
 RT intergenetic transfer of chromosomal genes between major human  
 RT pathogens".  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:12381-12385(1998).

DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE A/G-specific adenine glycosylase.  
GN NM01396.  
OS *Neisseria meningitidis* (serogroup B).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC58 / Serogroup B;  
RX MEDLINE=20175755; PubMed=10710307;  
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisten J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Groll M.L., Debroy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Ciecko A., Parisey D.S., Blair E., Clifton H., Clark E.B.,  
RA Colton M.D., Ulfelder T.R., Khouri H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Masiagnan V., Pizzo M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain  
MC58.";  
RL Science 287:1809-1815(2000).  
EMBL: AE002488; AAF41760.1; -.

DT 01-OCT-2000 (TremBLrel. 15, Created)  
DT 01-OCT-2000 (TremBLrel. 15, Last annotation update)  
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)  
DE Adenine glycosylase.  
GN BH0931.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=6665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fujii F., Hirata C., Nakamura Y., Ogawara N., Kohara S.,  
RA Horiuchi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL: AP001510; BAB04650.1; -  
DR HSSP: P17802; 1MIN  
DR InterPro: IPR004035; EndoIII\_FCL.  
DR InterPro: IPR003265; Endo\_3C.  
DR InterPro: IPR003651; Fes\_bind.  
DR InterPro: IPR000445; Hhh.





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 13:40:49 ; Search time 23 Seconds

(without alignments)  
664,508 Million cell updates/sec

Title: US-09-938-901-2  
Perfect score: 1700  
Sequence: 1 YEAMRKALLAWYRENAAPLP.....VLKRALPLAHNGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	484.5	28.5	350	1	MUTY_SALTY
2	471.5	27.7	350	1	MUTY_SALTY
3	445	26.2	378	1	MUTY_HAELIN
4	442.5	26.0	461	1	MYH_SCHPO
5	402	23.6	347	1	MUTY_BUCAP
6	378.5	22.3	350	1	MUTY_BUCAL
7	274.5	16.1	221	1	GMR_METTF
8	250.5	14.7	99	1	MUTY_ABRHY
9	227	13.4	213	1	END3_THEMA
10	199	11.7	279	1	UVEN_MICLU
11	176	10.4	259	1	NTHL_CABEL
12	172	10.1	219	1	END3_BACSU
13	171	10.1	344	1	Y613_METVA
14	164.5	9.7	312	1	NTHL_HUMAN
15	161.5	9.5	300	1	NTHL_MOUSE
16	158.5	9.3	355	1	END3_SCHPO
17	158	9.3	219	1	END3_SYNY3
18	145	8.5	211	1	END3_ECOLI
19	134	7.9	211	1	END3_HAELIN
20	131.5	7.7	245	1	END3_MYCTU
21	130.5	7.7	245	1	END3_MYCTU
22	115.5	6.8	211	1	END3_TREPA
23	111.5	6.6	212	1	END3_RICPR
24	111	6.5	1729	1	DME_ARATH
25	107	6.3	1505	1	CUT2_HUMAN
26	106	6.2	1309	1	DMLE_ARATH
27	101.5	6.0	775	1	PMIP_SCHPO
28	100.5	5.9	644	1	YAI4_TREPA
29	100	5.9	1132	1	TERT_HUMAN
30	97.5	5.7	209	1	END3_BUCAP
31	97.5	5.7	210	1	END3_BUCAL
32	97.5	5.7	432	1	Y41L_RHISN
33	97	5.7	525	1	NAB2_HUMAN

34	97	5.7	635	1	VP40_HSV11
35	93	5.5	309	1	FDHE_PSEAE
36	93	5.5	1067	1	MICA_HUMAN
37	92.5	5.4	345	1	MUTL_THETH
38	92	5.4	502	1	GAG_SIVGB
39	91.5	5.4	376	1	PROB_THETH
40	90	5.3	873	1	FPS_FUSV
41	89.5	5.3	380	1	MTG2_YEAST
42	89.5	5.3	462	1	TRPE_THETH
43	88.5	5.2	1103	1	KEIC_HUMAN
44	88	5.2	513	1	NOR1_ALCEB
45	88	5.2	525	1	NAB2_MOUSE

## ALIGNMENTS

RESULT 1  
MUTY\_SALTY STANDARD; PRT; 350 AA.  
ID ID  
AC 005869;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE A/G-specific adenine glycosylase (EC 3.2.2.-).  
GN MUTY OR MUTB OR STM3110.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_Taxid=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GM1803;  
RA MEDLINE=93123173; PubMed=8419300;  
RA Desiraju V., Shanabroch W.G., Lu A.L.;  
RT "Nucleotide sequence of the Salmonella typhimurium mutb gene, the  
RT homolog of Escherichia coli muty.";  
RL J. Bacteriol. 175:541-543(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SCS1412 / ATCC 700720;  
RA MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewel N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
CC -1- FUNCTION: ADENINE GLYCOSYLASE ACTIVE ON G-A AND C-A MISMATCHES.  
CC -1- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE  
CC CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER  
CC POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M86634; AAA27165.1; -  
CC EMBL: AE008843; AAL21985.1; -  
CC PIR: A40647; A40647.  
CC HSSP: P17802; 1MUN.  
CC ScyGene: SG10238; muty.  
CC InterPro: IPR003265; Endo\_3c.  
CC InterPro: IPR004035; EndoIII\_FCL.  
CC InterPro: IPR004036; EndoIII\_Hnh.  
CC InterPro: IPR003651; Fes\_bind.

```

DR InterPro: IPR000445; Hnh.
DR InterPro: IPR003583; Hnh.
DR InterPro: IPR005760; MutY.
DR Pfam: PF00730; Hnh-GPD; 1.
DR SMART: SM00478; ENDO3C; 1.
DR SMART: SM00525; PES; 1.
DR SMART: SM00278; Hnh1; 1.
DR TIGRfam: TIGR01084; mutY; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE: PS01155; ENDONUCLEASE_III_2; 1.
KW DNA repair; Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S;
FT METAL
FT METAL
FT METAL
FT METAL
SQ SEQUENCE 350 AA; 39409 MW; 5C55DB4D7B7BB69F CMC64;

Query Match
Best Local Similarity 28.5%; Score 484.5; DB 1; Length 350;
Matches 126; Conservative 58; Mismatches 115; Indels 49; Gaps 14;

QY 8 LAWYRENNAR-PLPRGKDPKRYLVSEVLLOOTRVEQALPYRRFLREPTLKALAAS 66
DB 10 VLDWYDKYGRKTLPMQIKTKPTKWLSEVMLOOTQTVIYFERRMARFPVVDLANAP 69
QY 67 LEEVRYWOGAGYTRRREHRLRLARSVE-----ELPPSAELRGIPGCPYTAANAASIA 121
DB 70 LDEVHLMTGLGYTAARNRLHKAQOAVATLHGSEFPOTFAETIALPGVGRSTAGAILSLA 129
QY 122 FGEYAAVAVGNSVRVLSRLFARES--PKEKE-----LPAAGILP-EGVDPCVWMOALMEL 175
DB 130 LGRHYPTLDGNVNRKYLARCAYVSGMPGKEVENTLMTLSEQVTPARGVER--ENQAMMDL 187
QY 176 GATVCLPKRPGACGL--GACRCKEAPGRYPAPRRKAKEELVALVILGKGVHLER 233
DB 188 GAAYCTRSKPKCITLCPILONGCIAAANEHSRYRPEKPKKQTLPTPTGYFTLL--OHQNE 243
QY 234 L-----EGRFQGLGVLPFRPE-ELPGREAAAGVSRILGAV--RHATTHRLRY-- 280
DB 244 IFLAQRPSPGLMGGLGYCFOPFARDELREMLAQHNADNLQOLNARFRTFSHFLLDLPV 303
QY 281 -----EVRGALMEGEGEDP---WKRPLPKLMKRYLRKALPL 313
DB 304 MWLPVSLDACMDGCSALMWNLNLAOPSPVGLAIAVERLLQO-LRTGAPV 350

RESULT 2
MUTY_ECOLI
AC P17802; STANDARD; PRT; 350 AA.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE A/G-specific adenine glycosylase (EC 3.2.2.-).
GN MUTY OR MICA OR A2961.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=90326504; PubMed=2197596;
RX "MutY, an adenine glycosylase active on G-A mispairs, has homology to
RT endonuclease III."
RL Nucleic Acids Res. 18:3841-3845(1990).
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RC MEDLINE=91161503; PubMed=2001994;

```

```

RA Tsai-Wu J.-J., Radiceella J.P., Lu A.-L.;
RT "Nucleotide sequence of the Escherichia coli mica gene required for
RL A/G-specific mismatch repair: identity of mica and mutY."
RN Bacteriol. 173:1902-1910(1991).
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (1.2 ANGSTROMS).
RA Guan Y., Manuel R.C., Atwal A.S., Parikh S.S., Mol C.D., Miller J.H.,
RA Lloyd S., Tainer J.A.;
RT "MutY catalytic core, mutant and bound adenine structures define
RL specificity for DNA repair enzyme superfamily."
CC Nat. Struct. Biol. 5:1058-1064(1998).
CC -1- FUNCTION: ADENINE GLYCOSYLASE ACTIVE ON G-A MISPAIRS. MUTY ALSO
CC TO THE OXIDATIVELY DAMAGED FORM OF GUANINE: 7,8-DIHYDRO-8-
CC OXOGUANINE.
CC -1- COPACATOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE
CC CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER
CC POSITIONING OF THE ENZYME ALONG THE DNA STRAND.
CC -1- SIMILARITY: BELONGS TO THE NTP/MUTY FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52391; CA36624.1; -
DR EMBL: M59471; AAA72957.1; -
DR EMBL: U28377; AAA69128.1; -
DR EMBL: AE000378; AAC75998.1; -
DR PIR: B38535; B38535; -
DR PDB: 1MUD; 29-OCT-99.
DR PDB: 1MUN; 26-AUG-99.
DR PDB: 1MUJ; 20-AUG-99.
DR PDB: 1KG2; 26-NOV-02.
DR PDB: 1KG3; 26-NOV-02.
DR PDB: 1KG4; 26-NOV-02.
DR PDB: 1KG5; 26-NOV-02.
DR PDB: 1KG6; 26-NOV-02.
DR PDB: 1KG7; 26-NOV-02.
DR PDB: 1KJ; 10-APR-02.
DR EcoGene: EG10627; mutY.
DR InterPro: IPR003365; Endo_3C.
DR InterPro: IPR004035; EndoIII_FCL.
DR InterPro: IPR004036; EndoIII_Hnh.
DR InterPro: IPR003651; Res_bind.
DR InterPro: IPR000445; Hnh.
DR InterPro: IPR003583; Hnh.
DR Pfam: PF00730; Hnh-GPD; 1.
DR Pfam: PF00653; Hnh.
DR SMART: SM00478; ENDO3C; 1.
DR SMART: SM00525; PES; 1.
DR SMART: SM00278; Hnh1; 1.
DR TIGRfam: TIGR01084; mutY; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE: PS01155; ENDONUCLEASE_III_2; 1.
KW DNA repair; Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S; 3D-structure;
FT METAL

```



```

FT METAL 199 199 IRON-SULFUR (4FE-4S).
FT METAL 202 202 IRON-SULFUR (4FE-4S).
FT METAL 208 208 IRON-SULFUR (4FE-4S).
FT HELIX 3 17
FT HELIX 23 25
FT HELIX 30 40
FT HELIX 41 41
FT HELIX 45 58
FT HELIX 62 67
FT HELIX 70 77
FT TURN 78 79
FT TURN 83 83
FT HELIX 84 99
FT TURN 101 102
FT HELIX 108 112
FT TURN 113 113
FT TURN 115 116
FT HELIX 119 130
FT HELIX 139 148
FT TURN 149 150
FT TURN 155 156
FT HELIX 158 171
FT TURN 175 176
FT HELIX 177 190
FT TURN 191 191
FT HELIX 193 201
FT TURN 203 207
FT HELIX 209 213
FT TURN 214 214
FT HELIX 216 218
SQ SEQUENCE 350 AA; 39149 MW; C7D3657C03EBB4F CRC64;

```

Query Match 27.7%; Score 471.5; DB 1; Length 350;

Best Local Similarity 35.1%; Pred. No. 1e-29; Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

```

QY 8 LLAMARENAR-PLPWGEKDPYRVLYSEVLLQOTRVQALPYRRFLERFPTLKALAAAS 66
DB 10 VLDMYDKYGRKTLPMQDKTPYKVMISEVWLQOTQVATVYPERFRFARFPTVLDLNAAP 69
QY 67 LEEVLRVWQAGAGYRRRAEHLRLARSVEEL-----PPSFAELRGLPGLPYTAANAVASIA 121
DB 70 LDEVLRHLMGTGLGYAARARNLHKAQOAVATLHGKFPETFEVVALPVGGRSTAGAILSLTS 129
QY 122 FGEERVAADVGNVRVLSRLFARES-PKEKE-----LEFLAGLLE-CYDGCWNAQALMEL 175
DB 130 LGAHFLIDGNVNRVLAARCTAVASGMPGKKEVENKMLSLSEQVTPAVGVER--FNQAMMDL 187
QY 176 GATVCLPKRPRGACAPL--GAFGRGKEAPGRYPAPRRRAKEERLVALVILGRKGVHL-- 231
DB 188 GANICTRSRKRKCSCLPLQNCICIAANNNSMALYPGKKRQTLPERGTGFLLLQHEDEVLLA 247
QY 232 -ERLEGRFOGLYVPLPEPPEE-----LPGREAAFGVRSRDLGEVRAHLTHRRRLV----- 280
DB 248 QRPSPGLMGLYCQCFQFADEESLRQWLAQRIADNLTO-LTAFRHTSFHFHLDIVPMWL 306
QY 281 -----EVRGALWEGEGEDP---WKRLPKLMEVLEKALPL 313
DB 307 PVSSFTGCMDEGNALWNLNAQPPSVGLAIAVERLLQO-LRTGAPV 350

```

# RESULT 3

MUTY\_HAEIN STANDARD; PRT; 378 AA.

AC P44320;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE A/G-specific adenine glycosylase (EC 3.2.2.-).

GN MUTY OR H10759.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus.

```

OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: ADENINE GLYCOSYLASE ACTIVE ON G-A AND C-A MISPAIRS
CC (BY SIMILARITY).
CC -1- COPACITOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE
CC CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER
CC POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32760; AAC22418.1; -.
DR PIR: C64091; C64091.
DR HSSP: P17802; IM0Y.
DR TIGR: H10759; -.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR004036; EndoIII_FCL.
DR InterPro: IPR003651; EndoIII_HNH.
DR InterPro: IPR000445; HNH.
DR InterPro: IPR005760; MutY.
DR Pfam: PF00730; HNH-GPD; 1.
DR Pfam: PF00633; HNH; 1.
DR SMART: SMO0478; ENDO3c; 1.
DR SMART: SMO0525; FES; 1.
DR TIGRFAMs: TIGR01084; muty; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE: PS01155; ENDONUCLEASE_III_2; 1.
DR DNA repair; Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S;
KW Complete proteome.
FT METAL 197 197 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 204 204 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 207 207 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 213 213 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 378 AA; 43505 MW; C443F62513182A21 CRC64;

```

Query Match 26.2%; Score 445; DB 1; Length 378;

Best Local Similarity 34.3%; Pred. No. 1.3e-27; Matches 111; Conservative 59; Mismatches 122; Indels 32; Gaps 11;

```

QY 6 KALLAWYRENAR-PLPWGEKDPYRVLYSEVLLQOTRVQALPYRRFLERFPTLKALAA 64
DB 13 KSVLAWYDKRGKRLPMQDKTPYKVMISEVWLQOTQVATVYPERFRFARFPTVLDLNAAP 72
QY 65 ASLEEVRVWQAGAGYRRRAEHLRLARSVEE-----LPPSFAELRGLPGLPYTAANAVAS 119
DB 73 ASQDEVLRHLMGTGLGYAARARNLHKAQOAVATVYPERFRFARFPTVLDLNAAP 132
QY 120 IAFGERVAADVGNVRVLSRLFA-----RSPKRELFAALQGLPGEVDPGVNAQALME 174
DB 133 SVNLQPPYLDGNVNRVLAARCTAVASGMPGKKEVENRMLALTEQVTP--TRVADENQAMMD 191

```

QY 175 LGATVCLPRKPRGACPLGAFRCGKEAP--RYPAPKRRAKRELVALLGRKG-VHL 231  
 Db 192 ICAMCMKTRKPKDCPLNTIDCLAVKYNKMKPKPKAMEKTYTFLITLSTKNGKVL 251  
 QY 233 ERLE--GREGGLYGVPLPPE---ELPGRAAFGVRSRLGVRHALTRRLRY----- 280  
 Db 232 EORENSGLMGGLFCFPPPEKSSLLHFLAQEKYTHQWEP--SFRRTFHFHLDIPIYA 309  
 QY 281 EVRGALWEGEGEDPWRKRLPLKME 304  
 Db 310 EMESTLCEQANLDMR---KVMK 329

## RESULT 4

MYH\_SCHPO

ID MYH\_SCHPO STANDARD; PRT; 461 AA.

AC Q10159; Q74679;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE A/G-specific adenine DNA glycosylase (EC 3.2.2.-).

GN MYH OR SPAC26A3.02.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OX NCBI\_taxid=4896;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, AND SUBUNIT.

RX MEDLINE=98411305; PubMed=9737967;

RA Lu A.-L., Fawcett W.P.;

RT "Characterization of the recombinant MutY homologue, an adenine DNA

glycosylase, from yeast Schizosaccharomyces pombe.";

RL J. Biol. Chem. 273:25098-25105(1998).

RP [2]

SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream K.A., Lyne M., Lyne R., Stewart A.,

RA Sgouras K., Peat N., Hayes J., Baker S., Baaham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

RA Gentles S., Gobie A., Hamlin N., Harris D., Hildalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volkheart G., Aert R., Robben J., Grymporter B.,

RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Filicz C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,

RA Dague R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,

RA Cerrito L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RA "The genome sequence of Schizosaccharomyces pombe.";

RL Nature 415:871-880(2002).

CC -1- FUNCTION: ADENINE GLYCOSYLASE ACTIVE ON G-A MISPAIRS.

CC CATALYTIC ACTIVITY, BUT WHICH IS NOT INVOLVED IN THE PROPER

CC POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).

CC -1- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE

CC SUBUNIT: Monomer.

CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF053340; AAC36307.1; --  
 DR EMBL: Z69240; CAA03225.1; --  
 DR PIR: T38390; T38390.  
 DR PIR: T43679; T43679.  
 DR HSP: P17802; IMON.  
 DR GeneDB: Spombe; SPAC26A3.02; --  
 DR InterPro: IPR003265; Endo\_3c.  
 DR InterPro: IPR004035; Endo\_1c.  
 DR InterPro: IPR004036; Endo\_1c.  
 DR InterPro: IPR003651; Fes\_bind.  
 DR InterPro: IPR000445; Hnh.  
 DR Pfam: PF00730; Hnh-GPD; 1.  
 DR Pfam: PF00633; Hnh; 1.  
 DR SMART: SM00478; ENDO3c; 1.  
 DR SMART: SM00525; PES; 1.  
 DR PROSITE: PS00764; ENDONUCLEASE\_III\_1; 1.  
 DR PROSITE: PS01155; ENDONUCLEASE\_III\_2; 1.  
 DR DNA repair: Hydrolyase; Glycosidase; Iron-sulfur; 4Fe-4S.  
 KW METAL.  
 FT METAL 226 226 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 233 233 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 236 236 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 242 242 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT CONFLICT 7 7 S -> F (TN REF. 1).  
 SQ SEQUENCE 461 AA; 52930 MM; 5908E86020E2242 CRC64;

Query Match 26.0%; Score 442.5; DB 1; Length 461;  
 Best Local Similarity 31.5%; Pred. No. 2,6e-27;

Matches 121; Conservative 52; Mismatches 94; Indels 117; Gaps 15;

QY 1 YEAMRRALLAWYRENAKPLPMR-----GEKDP-----YVLYSEVLLOOTRY 42  
 Db 18 VEFRESLIOFYDKTRILPMWKKECIPPSDSPLEDMEOQVORLYEVLVSEIMLOOTRY 77  
 QY 43 EQALPYRFFRFPPTKALAAAE-EVLRWQAGYRRAEHLRLARSVELPSPA 101  
 Db 78 EYVKRYTMMATLPLTSCAEAEVNTQVPLMSGMFGYRCKRLHQOCHLAKLPS-- 135  
 QY 102 EL-----RGLPLGPTYAAVAASIAFGRVAAGVNRVLSLFARESKENE-- 150  
 Db 136 ETPRTGEMAKGIGVCGPYTAGAVLSIAMKOPCIYDGNVIRVLSALAIHSDCSKGAN 195  
 QY 151 --LFALAGLLPEGVDPGVNMOALMEIGATVCLPKPRGACPLGAFCRG----- 198  
 Db 196 ALIMKLAEVLDYVRGDEDMQALMELGATTCPOSPPSCVSPISFICAKVQEQVINDG 254  
 QY 199 -----KEAD-----GRYPK-PRKRAKRELVALLGRKG-VHL 231  
 Db 235 NTKIYDIEDVCNICIDIPSKEDLQNMVWVAKRPVPAKTKREER--ALV----- 304  
 QY 232 ERLEGRGGLYGVPLPPELPGREAAFGVRSRLGVRHALTRRLRYEVRGALWEGE- 290  
 Db 305 -----IFOKIDPSTKEFKFLIRKP-----SAGILAGLWDEPPT 337  
 QY 291 ---GEDPWRKRLPLKMEKYLKAL 311  
 Db 338 IERQESW---PKMDAEFOKSI 357

## RESULT 5

MUTY\_BUCAP

ID MUTY\_BUCAP STANDARD; PRT; 347 AA.

AC O84926;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE A/G-specific adenine glycosylase (EC 3.2.2.-).

GN MUTY OR BUG534.

OS Buchnera aphidicola (subsp. Schizaphis graminum).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Buchnera.

OX NCBI\_TaxID=98794;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22084549; Pubmed=12089438;

RA Tamas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S., Wernegreen J.F., Sandstrom J.P., Moran N.A., Andersson S.G.E.;

RT "50 million years of genomic stasis in endosymbiotic bacteria.";

RL Science 296:2376-2379(2002).

CC -1- FUNCTION: Adenine glycosylase active on G-A and C-A mispairs (By similarity).

CC -1- COFACTOR: Binds a 4Fe-4S cluster which is not important for the catalytic activity, but which is probably involved in the proper positioning of the enzyme along the DNA strand (By similarity).

CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AE014127; AAM68075.1; -.

DR InterPro: IPR003265; Endo\_3c.

DR InterPro: IPR004035; EndoIII\_FCL.

DR InterPro: IPR004036; EndoIII\_HhH.

DR InterPro: IPR003651; Fes\_bind.

DR InterPro: IPR005760; MutY.

DR Pfam: PF00730; HhH-GPD; 1.

DR Pfam: PF00633; HhH; 1.

DR SMART: SM00478; ENDO3c; 1.

DR SMART: SM00525; FES; 1.

DR TIGRFSMS: TIGR01084; muty; 1.

DR PROSITE: PS00764; ENDONUCLEASE\_III\_1; 1.

DR PROSITE: PS01155; ENDONUCLEASE\_III\_2; FALSE\_NEG.

DR DNA repair: Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S;

KW Complete proteome.

FT METAL 192

FT METAL 199

FT METAL 202

FT METAL 208

SO SEQUENCE 347 AA; 41301 MW; D6F6279175293587 CRC64;

Query Match 23.6%; Score 402; DB 1; Length 347;

Best Local Similarity 34.2%; Pred. No. 2,7e-24;

Matches 77; Conservative 59; Mismatches 75; Indels 14; Gaps 5;

QY 1 VEAARKALLAWYRENAAP-LPWNGEKDPYRVLYSEVLLQOTRVEQALPYRRFLERPTL 59

DB 3 IYVFSLDIINWYHINRKNLPWKKDKTLTKVWSEIMLQOTYVKTALPYRKNFISRPNI 62

QY 60 KALAAASLEEVLRVWVGAGYRYRAEHLRLARSV-----ELPSPFAELNGLGIGPYTA 114

DB 63 QSLNOSKLDLIDLCWGLGTYKRAENITKVTIKKEPFGKPTGSDLLKLGIGSTA 122

QY 115 AAVASIFGERVAADVGNRVRLSRL-----FARESPEKEELPALAAGLLPEGVDGCVWN 169

DB 123 GALTSLSLDYFFILEGNVKNRILMRYGIIGYTEKKIEOKLMVLIETLPIH-NGSFSN 181

QY 170 QALMEIATVCLPKPRPGACPLGAFRCGKEAPG--RYPAARRR 212

DB 182 OGIMDIGALICTPKNKCMLCPLIQCIAYKERNWIKYPLKRRK 226

RESULT 6

MUTY\_BUCAI STANDARD; PRT; 350 AA.

ID AC P57617;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE A/G-specific adenine glycosylase (EC 3.2.2.-).

GN MUTY OR BU552.

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum

OS symbiotic bacterium).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Buchnera.

OX NCBI\_TaxID=118099;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Tokyo 1998;

RX MEDLINE=20445173; Pubmed=10993077;

RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;

RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. Aps.";

RL Nature 407:81-86(2000).

CC -1- FUNCTION: ADENINE GLYCOSYLASE ACTIVE ON G-A AND C-A MISPAIRS (BY SIMILARITY).

CC -1- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AP001119; BAB13244.1; -.

DR HSSP: P17802; IMY.

DR InterPro: IPR003265; Endo\_3c.

DR InterPro: IPR004035; EndoIII\_FCL.

DR InterPro: IPR004036; EndoIII\_HhH.

DR InterPro: IPR003651; Fes\_bind.

DR InterPro: IPR005760; MutY.

DR InterPro: IPR005760; MutY.

DR Pfam: PF00730; HhH-GPD; 1.

DR Pfam: PF00633; HhH; 1.

DR SMART: SM00478; ENDO3c; 1.

DR SMART: SM00525; FES; 1.

DR TIGRFSMS: TIGR01084; muty; 1.

DR PROSITE: PS00764; ENDONUCLEASE\_III\_1; 1.

DR PROSITE: PS01155; ENDONUCLEASE\_III\_2; 1.

DR DNA repair: Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S;

KW Complete proteome.

FT METAL 192

FT METAL 199

FT METAL 202

FT METAL 208

SO SEQUENCE 350 AA; 41589 MW; F4567EFEF9B92BD1 CRC64;

Query Match 22.3%; Score 378.5; DB 1; Length 350;

Best Local Similarity 26.6%; Pred. No. 1.9e-22;

Matches 91; Conservative 86; Mismatches 114; Indels 51; Gaps 12;

QY 8 LLAWYRENAAP-LPWNGEKDPYRVLYSEVLLQOTRVEQALPYRRFLERPTLKAALAA 66

DB 10 VLMWYHKNRKKDLPMQINKTLTYWSEIMLQOTTVASALPYRKFLLNPNIKSLNDSK 69

QY 67 LEEVLRVWOGAGYRYRAEHLRLARSVE-----LPPSFAELNGLGIGPYTA 121

DB 70 LDVLYLWMSGLGYNNRAKNIKYKSAQILKKRYKGIFPDGFSNIITQLPGIGSTAGAILSL 129

QY 122 FGERVAADVGNRVRLSRLA-----RESPEKEELPALAAGLLPEGVDGCVWN 176

DB 130 LNFEPYPLDGNVNRILTVRYGIGLKDKEKLMWIIESTLPIH-NTGKFNQGMMDIG 188

QY 177 ATVCLPKPRPGACPLGAFRC--RGKEAPGRYPAARRKRAEERLVALVLGG-----RK 227

```

Db      189 ASICISIKPKCTTOLPKKECIAIEKKWKEPYLKNIKTLPOKISWFIITIHENFWLKK 248
Qy      228 GVHLEREGRFQGYGVLPPEPELGRNACVSRPLG-----EVRHALTH 275
Db      249 NTEQE-----IMKELFCPEPKFKKE-----EALIMLEKKININTCENMISFPHKSHILH 300
Qy      276 RR-----LVEVRKALMGEGEDPW---KRP---LPTLKEVLT 307
Db      301 INPLRLRPISEFEFKENKKIMYLNKPNQHIGLRPRYQKITL 342

RESULT 7
ID      GTRM_METTF      STANDARD:      PRT:      221 AA.
AC      P29588:
DT      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DE      15-SEP-2003 (Rel. 42, Last annotation update)
OS      Possible G-T mismatches repair enzyme (Ec 3.2.2.-) (ORF10).
OG      Plasmid pPV1.
OC      Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC      Methanobacteriaceae; Methanothermobacter.
RN      NCBI_TaxID=145262;
[1]
RC      SEQUENCE FROM N.A.
RP      STRAIN=DSM 3848 / THF;
RA      MEDLINE=93126090; PubMed=1336177;
RX      Noelling J., van Eeden F.J.M., Egeen R.I.L., de Vos W.M.;
RT      "Modular organization of related Archaeal plasmids encoding different
RT      restriction-modification systems in Methanobacterium
RT      thermoformicum."
RL      Nucleic Acids Res. 20:6501-6507(1992).
RL      -1- FUNCTION: COULD ACT IN DNA REPAIR. RESTRICTION METHYLASE
CC      M METH1, WHICH IS ENCODED BY THIS PLASMID, GENERATES
CC      5-METHYLCTYOSINE WHICH IS, ESPECIALLY UNDER THERMOPHILIC
CC      CONDITIONS, SUBJECT TO DEAMINATION RESULTING IN G-T MISMATCHES.
CC      THIS PROTEIN COULD CORRECT THESE MISMATCHES.
CC      -1- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE
CC      CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER
CC      POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE NH/MUTY FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collabora-
CC      tion between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC      or send an email to license@isb-slb.ch).
CC      -----
CC      EMBL: X68366; CANA4833.1; -.
DR      PIR: S30312; S30312.
DR      PDB: 1KEA; 23-JAN-02.
DR      InterPro: IPR003265; Endo_3c.
DR      InterPro: IPR004035; EndoIII_FCL.
DR      InterPro: IPR004036; EndoIII_HNH.
DR      InterPro: IPR003651; Res_Bind.
DR      InterPro: IPR000445; HNH.
DR      Pfam: PF00730; Hnh-GPD; 1.
DR      Pfam: PF00633; HNH; 1.
DR      SMART: SM00478; ENDO3c; 1.
DR      SMART: SM00525; FES; 1.
DR      PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
DR      PROSITE: PS01155; ENDONUCLEASE_III_2; 1.
DR      DNA repair: Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S; Plasmid;
DR      Hypothetical protein; 3d-structure.
FT      METAL      197
FT      METAL      197
FT      METAL      204
FT      METAL      207
FT      METAL      213
FT      METAL      213
SQ      SEQUENCE      221 AA; 2540 MW; D085FCECEIDDC4B62 CRC64;

```

Query Match	16.18	Score 274.5	DB 1	Length 221
Best Local Similarity	31.7%	Pred. No. 1.5e-14		
Matches	63	Conservative	40	Mismatches 85; Indels 11; Gaps 4
QY	8	LLAWYREALPPLMPGCEGDPYRIYSEVILLQOTRVEQALPYRRFLTEREPTLKALAAASL	67	
Db	16	ILTFEMNTDRDFPFRHRTROPYILITELTLNRTTAGVKKRIYDKFFVKKCEFDILKTPK	75	
QY	68	EEVLRVQMGQGYTRAEHLRLRLARSV-----ELPSFAELRGLDPLGDPYAAVAASIA	121	
Db	76	SEIAKDKEIGELSNQAEQELKELARVINDYGVGRPNRRAIIDLDGCVKTYCAAVMKLA	135	
QY	122	FGERVAAVDGNAVRYVLSRLFARE-----SPKKELEFALAQGLPEGVDPGVNQAALMELGA	177	
Db	136	FGKKAAMVDANFVRIYINKRYFGGSTEINLYNHKKALMELAEITLVPGGCKRD-FVTLGLMPSA	194	
QY	178	TYCLPRRPGCGACPLGAFNC	196	
Db	195	IICAPRRKPCCKCGMSKLC	213	

RESULT 8  
MUTY\_AERH

ID	MUTY_AERHY	STANDARD;	PRT;	99 AA.
AC	P46230;			

DT	01-NOV-1995	(Rel. 32, Created)
DT	01-NOV-1995	(Rel. 32, 1994-2000)

DT	01-NOV-1995 (Rel. 32, Last annotation update)
DE	A/G-specific adenine clusters

GN MUTY. *Neisseria meningitidis* glycosylase (EC 3.2.2.-) (Fragment)  
OS *Aeromonas hydrophila*

0C Bacteria; proteobacteria; Gammaproteobacteria; Aeromonadales  
0C Aeromonadaceae; *Aeromonas*

NCBI\_TaxID=644;  
[1]

SEQUENCE FROM N.A.

STRAIN=Ah65;  
MEDLINE=95050248; PubMed=7961440;

Canagildar R., Howard S.P.,  
RT "Isolation and characterization of a second and

RL extracellular protein secretion in *Aeromonas hydrophila*. " J. Bacteriol. 176:6819-6826(1994)

```

-!- FUNCTION: ADENINE GLYCOSYLASE ACTIVE ON G-A AND C-A MISPAIRS
    (BY SIMILARITY).
CC
CC

```

-1- COFACITOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE CATALYTIC ACTIVITY BUT WHICH IS IMPORTANT FOR THE STABILITY OF THE

```

-|- SIMILARITY. DEFINES NO OTHER POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY)
-|- SIMILARITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER
CCCCC
```

-----  
 BELONGS TO THE NTH/MUTY FAMILY.  
 -----  
 THIS SWISS-DBOM

between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. It is produced through a collaboration

use by non-profit institutions as long as its content is not modified and not used for commercial purposes.

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ish-eth.ch/en/and>).

or send an email to [license@slb.ch](mailto:license@slb.ch).

EMBL:	X81473;	NOT_ANNOTATED_CDS.
R		
HSSP:	P17802;	1MUN.

InterPro; IPR003265; Endo\_3c.  
InterPro; IPR004035; EndoIII FCL.

InterPro; IPR004036; EndoIII\_HhH.  
Pfam; PF00730; HhH-GPD; 1

PROSITE: PS00764; ENDONUCLEASE\_III\_1; PARTIAL.  
PROSITE: PS01155; ENDONUCLEASE\_III\_2; PARTIAL.

DNA repair; Hydrolyase; Glycosidase; Iron-sulfur; 4Fe-4S.  
 NON TER 99 00

SEQUENCE 99 AA; 11692 MW; 0806BDC67E1B0CB6 CRC64;

```
Query Match      14.78; Score 250.5; DB 1; Length 99.
Test Local Similarity
```

Matches	45;	Conservative	19;	Mismatches	22;	Indels	1.
Local similarity	51.7%;	Pred. No.	4.4e-13;				

1. The first step is to identify the problem or question that needs to be addressed. This involves understanding the context and the specific requirements of the task.

OY 8 LLMAYRENAR-PLPMWKEKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAAS 66  
 DB 13 ILDMYQJHGKRTLPYWOODKTPYRWYSEIMLQOQVATVPYQRFARFPDQALQAP 72  
 OY 67 LEEVLRWOGAGYRRARHLRLARV 93  
 DB 73 IDEVLHMTGLGYARARNLHKAQOI 99

RESULT 9  
 END3\_THEME STANDARD: PRT: 213 AA.  
 AC O9WYKO;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Endonuclease III (EC 4.2.99.18) (DNA-(apurinic or apyrimidinic site) lyase).  
 GN NTH OR TW0366.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogae; Thermotogaceae; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=MSB8 / DSM 3109;  
 MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RA "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima."  
 RT Nature 399:323-329(1999).  
 RL Nature 399:323-329(1999).  
 CC -1- FUNCTION: HAS BOTH AN APURINIC AND/OR APYRIMIDINIC ENDONUCLEASE ACTIVITY AND A DNA N-GLYCOSYLASE ACTIVITY. INCISES DAMAGED DNA AT CYTOSINES, THYMINES AND GUANINES. ACTS ON A DAMAGED STRAND, 5' FROM THE DAMAGED SITE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or apyrimidinic site in DNA is broken by a beta-elimination reaction, leaving a 3'-terminal unsaturated sugar and a product with a terminal 5'-phosphate.  
 CC -1- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AE001716; AAD35453.1;  
 DR PIR: F72387; F72387.  
 DR HSSP: P20625; ZABK.  
 DR TIGR: TM0366;  
 DR InterPro: IPR003265; Endo\_3c.  
 DR InterPro: IPR004035; EndoIII\_FCL.  
 DR InterPro: IPR004036; EndoIII\_HhH.  
 DR InterPro: IPR003651; Fes\_Dind.  
 DR InterPro: IPR000445; HhH.  
 DR InterPro: IPR003583; HhH\_1.  
 DR InterPro: IPR005759; Nth.  
 DR Pfam: PF00730; HhH-GPD; 1.  
 DR Pfam: PF00633; HhH; 1.  
 DR SMART: SM00478; ENDO3c; 1.  
 DR SMART: SM00525; FES; 1.  
 DR SMART: SM00278; HhH1; 2.  
 DR TIGRFAMs: TIGR01083; nth; 1.

DR PROSITE: PS00764; ENDONUCLEASE\_III\_1; 1.  
 DR PROSITE: PS01155; ENDONUCLEASE\_III\_2; 1.  
 KW Hydrolyase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;  
 KM Glycosidase; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.  
 FT METAL 180  
 FT METAL 187  
 FT METAL 187  
 FT METAL 190  
 FT METAL 196  
 SO SEQUENCE 213 AA; 24928 MM; D9716CD6FEF85D4 CRC64;

Query Match 13.4%; Score 227; DB 1; Length 213;  
 Best Local Similarity 34.4%; Pred. No. 7.3e-11;  
 Matches 66; Conservative 33; Mismatches 71; Indels 22; Gaps 7;

OY 24 EKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAASLEEVLRWOGAGYRR-R 82  
 DB 19 ETDFRVLISVLISQRTDENTERASKLEEVYTPQELAKAKEDLYDLKESGMROK 78  
 OY 83 AEHLRLARVSE-----LPPSFAELRGLPGLGPTAAVAASIAFGERRVAADGNVRVL 137  
 DB 79 AERIVEISRLIVEYGGRRVPSLELLKLPVGGKRTANIVLVGFKKPALAVDTHVRIS 138  
 OY 138 SRL--FARESPEKELEFALAOGILLPEGVPCW--NQALMELGATVCLPRPRGACPL 192  
 DB 139 NRLGWVKTPTPEETE--EALKLLPE---DLMGPIGSMVEFGRRICKPONPLCEECFL 192  
 OY 193 GAFG-----RGK 199  
 DB 193 KNHCEFYRRRK 204

RESULT 10  
 UVEN\_MICLU STANDARD: PRT: 279 AA.  
 ID UVEN\_MICLU  
 AC P46303;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ultraviolet N-glycosylase/Ar lyase (UV-endonuclease) (Pyrimidine dimer glycosylase).  
 GN PDG.  
 OS Micrococcus luteus (Micrococcus lysodeikticus).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 CC Micrococcales; Micrococcaceae; Micrococcus.  
 OX NCBI\_TaxID=1270;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-35.  
 RC STRAIN=ATCC 4698;  
 RX MEDLINE=96007490; PubMed=759510;  
 RA Pierson C.E., Prince M.A., Augustine M.L., Dodson M.L., Lloyd R.S.;  
 RT "Purification and cloning of Micrococcus luteus ultraviolet endonuclease, an N-glycosylase/abasic lyase that proceeds via an imino enzyme-DNA intermediate."  
 RT J. Biol. Chem. 270:23475-23484(1995).  
 RL J. Biol. Chem. 270:23475-23484(1995).  
 CC -1- FUNCTION: HAS BOTH AN APURINIC AND/OR APYRIMIDINIC ENDONUCLEASE ACTIVITY AND A DNA N-GLYCOSYLASE ACTIVITY. INITIATES REPAIR AT CTS-SYN PYRIMIDINE DIMERS. PROCEEDS VIA AN IMINO ENZYME-DNA INTERMEDIATE.  
 CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR UAG OCCURS BETWEEN CODONS FOR GLY-268 AND ALA-270. TWO FORMS OF 31 KDA AND 32 KDA HAVE BEEN DETECTED.  
 CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: U22181; AAA86508.1; ALT\_TERM.  
 DR HSSP: P20625; ZABK.

```
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR004035; EndoIII_PCL.
DR InterPro: IPR004036; EndoIII_PCL.
DR InterPro: IPR004651; Fes_bind.
DR InterPro: IPR000445; Hhh.
DR InterPro: IPR003583; Hhh_1.
DR Pfam: PF00730; Hhh-GPD; 1.
DR Pfam: PF00633; Hhh; 1.
DR SMART: SM00478; ENDO3c; 1.
DR SMART: SM00525; FES; 1.
DR TIGRfam: TIGR01083; nfh; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE: PS01155; ENDONUCLEASE_III_2; 1.
DR HydroLase: Nuclease; DNA repair; Glycosidase;
KW Iron-sulfur; 4Fe-4S.
FT CHAIN 1 268 UV ENDONUCLEASE 31 kDa FORM.
FT METAL 203 279 IRON-SULFUR (4FE-4S) (PROBABILE).
FT METAL 210 210 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 213 213 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 219 219 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 279 AA; 30469 MW; C385369A1827C005 CRC64;

Query Match
Best Local Similarity 11.7%; Score 199; DB 1; Length 279;
Matches 70; Conservative 31; Mismatches 95; Indels 20; Gaps 8;

QY 27 PYRLVSEVLLQOT---RVEQALPYRRFLERPTLKALAAASLEEVLRVMOGAGYR-R 82
DB 44 PFEILLVTLVSAQTDDVRVNAAP---ALFAHPPPAHMAAATEPELDELVRSGFTRNK 100
QY 83 AEHLRLARSV-----ELPPSPFAELRLGPGPYTAAVAASIAFGERYAAVDGNVRRVL 137
DB 101 ASAILRLSQELVGRHDEGVPARLEDVLPVGRKTAIVYLGNAFGPGITVTHGRILA 160
QY 138 SRL-FARESPKREKELFALAGLPEGVDPGVNQAIMELGATVYCLPRPRCGACPLGAC 196
DB 161 RRLGFETDEDPKGR-ARRGRVPVPRARMTWLSRLLFTHGRVCHARRPACGRCPIARNC 219
QY 197 ---RQKEAPGRYPAPRRRAK---EERLVALVLGR 226
DB 220 PSYAGETDPERARALLAVELKPGRELLLELLRAGR 255

RESULT 11
NTH1_CABEL
AC PS4137 STANDARD; PRT; 259 AA.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable endonuclease III homolog (EC 4.2.99.18) (DNA-(apurinic or
aprimidinic site) lyase).
GN R10E4.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Ainscough R.;
RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Has both an apurinic and/or aprimidinic endonuclease
CC activity and a DNA N-glycosylase activity (by similarity).
CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC aprimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -1- COFACTOR: Binds a 4Fe-4S cluster which is not important for the
CC catalytic activity, but which is probably involved in the proper
```

```
CC positioning of the enzyme along the DNA strand (By similarity).
CC -1- SIMILARITY: BELONGS TO THE NTH/MUVY FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC or send an email to license@sb.sib.ch.
CC EMBL: 250874; CA90766.1; -.
DR PIR: T24131; T24131.
DR HSSP: P20625; ZABK.
DR WormPep: R10E4.5; CE03559.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR004035; EndoIII_PCL.
DR InterPro: IPR004036; EndoIII_PCL.
DR InterPro: IPR003651; Fes_bind.
DR Pfam: PF00730; Hhh-GPD; 1.
DR Pfam: PF00633; Hhh; 1.
DR SMART: SM00478; ENDO3c; 1.
DR SMART: SM00525; FES; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE: PS01155; ENDONUCLEASE_III_2; 1.
KW Hypothetical protein; HydroLase; Nuclease; Endonuclease;
KW Multifunctional enzyme; DNA repair; Glycosidase; Lyase; Iron-sulfur;
KW 4Fe-4S.
FT METAL 187 187 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 194 194 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 197 197 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 203 203 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 259 AA; 28841 MW; F52E57810571EDBC CRC64;

Query Match
Best Local Similarity 10.4%; Score 176; DB 1; Length 259;
Matches 58; Conservative 33; Mismatches 70; Indels 20; Gaps 8;

QY 28 YRLVSEVLLQOTRVEQALPYRRFLERPTLKALAAASLEEVLRVMOGAGYR-RRAEHL 86
DB 29 FOVLVALMLSSQTDENVAAAMKRLKDHCIGLSTIGKILFEKVPDLETILCPGVFYRKAVYL 88
QY 87 HRLARSVE-----ELPPSPFAELRLGPGPYTAAVAASIAFGERYAAVDGNVRRVL 140
DB 89 OKTAKILKDDPSGDIPLDGLCALPVGPKMANLVMQIANGECGIAVDTHVHRISNRL 148
QY 141 -FARESPEKREKELFALAGLPEGVDPGVN---NQAIMELGATVYCLPRPRCGACPLGAC 196
DB 149 GWITSTPEKTOKAL-EILLPK-----SEWQPINHLVGFQMGQCPVPRPCGTC---LC 199
QY 197 R 197
DB 200 R 200

RESULT 12
END3_BACSU
AC P39788 STANDARD; PRT; 219 AA.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable endonuclease III (EC 4.2.99.18) (DNA-(apurinic or
aprimidinic site) lyase).
GN NTH OR JOOB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95219085; PubMed=7704260;
```

RA Bruand C., Sorokin A., Serror P., Ehrlich S.D.;  
 RT "Nucleotide sequence of the *Bacillus subtilis* *dnad* gene.";  
 RL Microbiology 141:321-322(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / Marburg;  
 RX MEDLINE=96349105; PubMed=8760912;  
 RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,  
 Serror P.;  
 RT "Sequence analysis of the *Bacillus subtilis* chromosome region between  
 the *serA* and *kgf* loci cloned in a yeast artificial chromosome.";  
 RL Microbiology 142:2005-2016(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Bertoletti M.G., Bessieres P., Bolotin A., Borchert S.,  
 Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 Guisepi G., Guy B.J., Haga K., Haeleth J., Harwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kasahara Y., Kieffer-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kunano M.,  
 Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 Parro V., Pohl T.M., Portetalle D., Potworlik S., Prescott A.M.,  
 Prescan E., Puig P., Purrelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Rocha E., Roche E., Rose M., Sadale Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
 Takenuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretli A.,  
 Viari A., Wambolt R., Wedler E., Wedler H., Weltenegeger T.,  
 Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.";  
 RL Nature 390:249-256(1997).  
 CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or  
 apyrimidinic site in DNA is broken by a beta-elimination reaction,  
 leaving a 3'-terminal unsaturated sugar and a product with a  
 terminal 5'-phosphate.  
 CC -1- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE  
 CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER  
 POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U11289; AAA80005.1; -;  
 DR EMBL: L47709; AAB38457.1; -;  
 DR EMBL: 299115; CAB14150.1; -;  
 DR PIR: I40525; I40525.  
 DR HSSP: P20625; 2ABK.  
 DR Subtilist; BG10956; nth.  
 DR InterPro; IPR003265; Endo\_3c.  
 DR InterPro; IPR004036; EndoIII\_FCL.  
 DR InterPro; IPR004036; EndoIII\_Hhh.

DR InterPro; IPR003651; Fes\_bind.  
 DR InterPro; IPR000445; HHH.  
 DR InterPro; IPR003583; HHH\_1.  
 DR InterPro; IPR005759; Nch.  
 DR Pfam; PF00730; Hhh-GPD; 1.  
 DR Pfam; PF00633; Hhh; 1.  
 DR SMART; SM00478; ENDO3c; 1.  
 DR SMART; SM00525; FES; 1.  
 DR SMART; SM00278; Hhh1; 1.  
 DR TIGRfam; TIGR01083; nth; 1.  
 DR PROSITE; PS00764; ENDONUCLEASE\_III\_1; 1.  
 DR PROSITE; PS01155; ENDONUCLEASE\_III\_2; 1.  
 KW Hydrolyase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;  
 KW Glycosidase; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.  
 FT METAL 189 189 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 196 196 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 199 199 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 205 205 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 SQ SEQUENCE 219 AA; 25000 MW; C38C66DA7948BA04 CRC64;  
 Query Match 10.18; Score 172; DB 1; Length 219;  
 Best Local Similarity 28.18; Pred. No. 1.5e-06;  
 Matches 56; Conservative 30; Mismatches 65; Indels 48; Gaps 6;  
 QY 26 DPYRVLYSEVLQO-----TRVEQALPYRRFLERPTIKALAAALEEYLRWQAGAY 79  
 Db 29 NPPELYVAVALASQCDALNRYTKL-----FQKKREDYIAVLEELQDQKISGL 82  
 QY 80 YR-RAEHLRLARVSV-----EELPPSFALRGGLGCPYTAAYASTAFGERVAADGVN 133  
 Db 83 YRNKAKNIQKLSKMIIDYGEVDRDELVKLPVGRKTAANYVSAVGFAPALAVDTYH 142  
 QY 134 RYVLSRL-----FARSPKEKELFALAGLLPVEDVPGVNNQALMEIGAT 178  
 Db 143 ERYSKRIGRCGRWDSEVLEKTKRKVPKE-----DMSVTHHRLIFGRY 187  
 QY 179 VCLPRPGCACPAGFCR 197  
 Db 188 HCKAQSPRCACPELISCR 206  
 RESULT 13  
 Y613\_METJA STANDARD; PRT; 344 AA.  
 ID Y613\_METJA  
 AC 058030;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative endonuclease MJ0613.  
 GN MJ0613.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 Sutton G.G., Blake J., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 Scott J.L., Geoghagen N.S.M., Meldrum J.F., Fuhmann J.L., Nguyen D.,  
 Uiterback T.R., Kelley J.M., Peterson J.D., Sadow P.M., Hanna M.C.,  
 Klenk H.-P., Roberts R.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 R. Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.";  
 RL Science 273:1058-1073(1996).  
 CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U67509: AAB98606.1; -  
 DR HSSP: P20625; ZABK.  
 DR TIGR: M0613; -

DR InterPro: IPR002837; DUF123.  
 DR InterPro: IPR003265; Endo\_3c.  
 DR InterPro: IPR004035; EndoII\_FCL.  
 DR InterPro: IPR004036; EndoIII\_HH.  
 DR InterPro: IPR003651; Fes\_bind.  
 DR InterPro: IPR000445; HH.  
 DR InterPro: IPR003583; HH\_1.  
 DR InterPro: IPR003583; HH\_1.  
 DR Pfam: PF001986; DUF123; 1.  
 DR Pfam: PF00730; HNH-GPD; 1.  
 DR Pfam: PF00633; HNH; 1.  
 DR ProDom: PD015467; DUF123; 1.  
 DR SMART: SM00478; ENDO3c; 1.  
 DR SMART: SM00525; FES; 1.  
 DR SMART: SM00465; GYC; 1.  
 DR SMART: SM00278; HNH; 1.  
 DR PROSITE: PS00764; ENDONUCLEASE\_III\_1; FALSE\_NEG.  
 DR PROSITE: PS01155; ENDONUCLEASE\_III\_2; 1.  
 DR Hypothetical protein: Hydrolase; Glycosylase; Iron-sulfur; 4Fe-4S;  
 KW Complete proteome.  
 FT METAL 184 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 190 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 193 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 201 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 SQ SEQUENCE 344 AA; 40567 MW; BC3BEFE1DE778B85 CRC64;

Query Match 10.1%; Score 171; DB 1; Length 344;  
 Best Local Similarity 28.4%; Pred. No. 3e-06;  
 Matches 55; Conservative 41; Mismatches 80; Indels 18; Gaps 7;

QY 8 LLAWRNENARPLPMNGEDPRVAVSEVLLQOTRVEQALPYRRFLERPPTLKALAASTL 67  
 DB 7 LKLRKNAVYETLAKDPRKVLITISARTKDEVTVEBSKLFKEKDDVLLNIDE 66  
 QY 68 EEVLKRVMOGAGYR-RAEHLRLARSVEE-----LPPSAELRGLPGLPPTAAVASIA 121  
 DB 67 EKLADLYPAGFYFNKAKNLKILKENYKNKVPDSLEELKLPGLGVGRKTAANLVITLA 126  
 QY 122 FGERVAADVGNVRVLSR--LFARSPKEKEFLAALQGLPBGVDPGVW---NOALMELG 176  
 DB 127 FNKKGICVDTHVHRICNRMEIVDTETPEETE-FELRKILPKK-----YMKVINLLVYFG 180  
 QY 177 ATYCLPRPRGCGAC 190  
 DB 181 REICSSK-SKCDKC 193

RESULT 14  
 NTH1\_HUMAN  
 ID NTH1\_HUMAN STANDARD: PRT; 312 AA.  
 AC P78549; Q99566; Q99784; Q998X2;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Endonuclease III-like protein 1 (EC 4.2.99.18).  
 GN NTH1 OR NTH1 OR OCTS3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RP MEDLINE=97144402; PubMed=8990169;  
 RA Asplund R., Rothwell D.G., Roldan-Arjona T., Anselmino C.,

RA Ward C.J., Cheadle J.P., Sampson J.R., Lindahl T., Harris P.C.,  
 RA Hickson I.D.,  
 RT "Cloning and characterization of a functional human homolog of  
 RT Escherichia coli endonuclease III.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 94:109-114(1997).  
 RN [2]

RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99051340; PubMed=9831664;  
 RA Imai K., Sarker A.H., Akiyama K., Ikeda S., Yao M., Tsutsui K.,  
 RA Shomori T., Seki S.,  
 RT "Genomic structure and sequence of a human homologue (NTH1/NTH1) of  
 RT Escherichia coli endonuclease III with those of the adjacent parts of  
 RT gene 222:287-295(1998).  
 RN [3]

RP SEQUENCE FROM N.A.  
 RA Rieck D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,  
 RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,  
 RA Goodwin L., Bryant J., Tesmer J., Melnick L., Longmire J., White S.,  
 RA Ueng S., Tatum O., Campbell C., Fawcett J., Maltbie M., Mista M.,  
 RA Deaven L.,  
 RT "Sequencing of human chromosome 16p13.3.";  
 RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]

RP SEQUENCE FROM N.A.  
 RA Rieder M.J., Braun A.C., Montoya M.A., Chung M.-W., Nguyen C.P.,  
 RA Nguyen D.A., Livingston R.J., Poel C.L., Robertson P.D.,  
 RA Schackwitz W.S., Sherwood J.K., Wiltrak L.A., Nickerson D.A.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 RN [5]

RP SEQUENCE OF 6-312 FROM N.A., FUNCTION, ACTIVE SITE, AND MUTAGENESIS OF  
 RP LYS-220.  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=98370989; PubMed=9705289;  
 RA Ikeda S., Biswas T., Roy R., Izumi T., Bologh I., Kurosky A.,  
 RA Sarker A.H., Seki S., Mitra S.,  
 RT "Purification and characterization of human NTH1, a homolog of  
 RT Escherichia coli endonuclease III. Direct identification of Lys-212 as  
 RT the active nucleophilic residue.";  
 RL J. Biol. Chem. 273:21585-21593(1998).  
 RN [6]

RP SEQUENCE OF 7-312 FROM N.A., AND FUNCTION.  
 RC TISSUE=Spleen;  
 RX MEDLINE=97197855; PubMed=9045706;  
 RA Hilbert T.P., Chang W., Boorstein R.J., Cunningham R.P., Teohor G.W.;  
 RT "Cloning and expression of the cDNA encoding the human homologue of  
 RT the DNA repair enzyme, Escherichia coli endonuclease III.";  
 RL J. Biol. Chem. 272:6733-6740(1997).  
 RN [7]

RP SEQUENCE OF 8-312 FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Pelngold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Dlatshenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Brownstein M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Hong L.,  
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson J.P., Prange C.,  
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs S.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Kizyewski M.I., Skalska U., Smalls D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).



RN [8] SEQUENCE OF 9-312 FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND  
 RP DEVELOPMENTAL STAGE.  
 RX MEDLINE-20342089; PubMed-10882850;  
 RA Luna L., Bjoras M., Hoff E., Rognes T., Seeberg E.;  
 RT "Cell-cycle regulation, intracellular sorting and induced  
 overexpression of the human NTH1 DNA glycosylase involved in removal  
 of formidopyrimidine residues from DNA.";  
 RL Mutat. Res. 460:95-104(2000).  
 RN SUBCELLULAR LOCATION.  
 RP MEDLINE-22419222; PubMed-12531031;  
 RA Ikeda S., Kohmoto T., Tabata R., Seki Y.;  
 RT "Differential intracellular localization of the human and mouse  
 endonuclease III homologs and analysis of the sorting signals.";  
 RL DNA Repair 1:847-854(2002).  
 CC -1- FUNCTION: Has both an apurinic and/or apyrimidinic endonuclease  
 activity and a DNA N-glycosylase activity. Incises damaged DNA at  
 cytosines, thymines and guanines. Acts on a damaged strand, 5'  
 from the damaged site. Required for the repair of both oxidative  
 DNA damage and spontaneous mutagenic lesions.  
 CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or  
 apyrimidinic site in DNA is broken by a beta-elimination reaction,  
 leaving a 3'-terminal unsaturated sugar and a product with a  
 terminal 5'-phosphate.  
 CC -1- COFACTOR: Binds a 4Fe-4S cluster which is not important for the  
 catalytic activity, but which is probably involved in the proper  
 positioning of the enzyme along the DNA strand.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Widely expressed with highest levels in heart  
 and lowest levels in lung and liver.  
 CC -1- DEVELOPMENTAL STAGE: Expression levels are regulated during the  
 cell cycle with increased levels during early and mid S-  
 phase.  
 CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.  
 CC -1- CAUTION: It is uncertain whether Met-1, Met-9 or Met-16 is the  
 initiator.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC  
 CC EMBL: U79718; AAB41534.1; -  
 DR EMBL: AB014460; BAA32695.1; ALT\_INIT.  
 DR EMBL: AC005600; AAC34209.1; -  
 DR EMBL: AF498098; AAM11786.1; -  
 DR EMBL: AB001575; BAA19413.1; ALT\_INIT.  
 DR EMBL: U81285; AAC51136.1; ALT\_INIT.  
 DR EMBL: BC003014; AAH03014.1; ALT\_INIT.  
 DR EMBL: BC000391; AAH00391.1; ALT\_INIT.  
 DR EMBL: Y09687; CAA70865.1; -  
 DR GeneW; HGNC:8028; NTHL1.  
 DR MIM: 602656; -  
 DR GO: GO:0005634; C:nucleus; IDA.  
 DR GO: GO:0019104; F:DNA N-glycosylase activity; IDA.  
 DR GO: GO:0004519; F:endonuclease activity; TAS.  
 DR GO: GO:0006284; P:base-excision repair; TAS.  
 DR GO: GO:0006296; P:nucleotide-excision repair, DNA incision, 5...; IDA.  
 DR InterPro: IPR003265; Endo3c.  
 DR InterPro: IPR004036; EndoIII\_Hhm.  
 DR InterPro: IPR003651; Fes\_Bldn.  
 DR InterPro: IPR000445; Hhm.  
 DR Pfam: PF00730; Hhm-GPD; 1.  
 DR Pfam: PF00633; Hhm-1.  
 DR SMART: SM00478; ENDO3c; 1.  
 DR SMART: SM00525; FES; 1.  
 DR PROSITE: PS00764; ENDONUCLEASE\_III\_1; FALSE\_NEG.  
 DR PROSITE: PS01155; ENDONUCLEASE\_III\_2; 1.  
 KW Hydrolase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;

KW Glycosidase; Lyase; Iron-sulfur; 4Fe-4S; Nuclear protein.  
 FT ACT\_SITE 220 220 NUCLEOPHILE IN THE N-GLYCOSYLASE  
 FT REACTION.  
 FT METAL 290 290 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 297 297 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 300 300 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 306 306 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT MUTAGEN 220 220 K->O: INACTIVATES ENZYME.  
 FT MUTAGEN 220 220 K->R: 85-FOLD REDUCTION IN ACTIVITY.  
 FT CONFLICT 9 10 MT -> TS (IN REF. 8).  
 FT CONFLICT 78 78 MISSING (IN REF. 8).  
 FT CONFLICT 151 151 M -> I (IN REF. 1).  
 FT CONFLICT 160 160 T -> A (IN REF. 1).  
 SQ SEQUENCE 312 AA; 34389 MW; 379816A1E0B4505050 CRC64;  
 Query Match 9.7%; Score 164.5; DB 1; Length 312;  
 Best Local Similarity 28.1%; Pred. No. 8.66-06;  
 Matches 56; Conservative 38; Mismatches 80; Indels 25; Gaps 7;  
 QY 12 YRENARPLPGRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAASLEVL 71  
 Db 119 YDSSAPFKVRR-----YQVLISLMSQTKDQVYAGAMQRLRAGLVDSITQDDMTLG 173  
 QY 72 RVWGAGCYRR-RAEHLRLARSYE----ELPSFAELRGIPGIGPYTAAVASIARGE- 124  
 Db 174 KLIVPGFMRSKYKVIKQTSALIQHYGSDIPASVAELVALPGVPMKHLAMAVMGTV 233  
 QY 125 RVAADVQNRVRL-----FARSPKEKELFALAGLPEGVPGW---NQLMELGA 177  
 Db 234 SGIAVDHVRHRIARLMTKATKSPBE-----TRAALBWLPRELHMEINGLLVGFQ 287  
 QY 178 TVCLPKRRPGACPLGAF 196  
 Db 288 QTLCPVPRCHACLNQALC 306  
 RESULT 15  
 NTH1\_MOUSE  
 ID NTH1\_MOUSE STANDARD; PRT; 300 AA.  
 AC O35980;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endonuclease III-like protein 1 (EC 4.2.99.18).  
 GN NTHL1 OR NTHL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN-BALB/C;  
 RX MEDLINE-98417643; PubMed-9743625;  
 RA Sarker A.H., Ikeda S., Nakano H., Terato H., Ide H., Imai K.,  
 RA Akiyama K., Tsutsui K., Bo Z., Kubo K., Yamamoto K., Yasui A.,  
 RA Yoshida M.C., Seki S.;  
 RT "Cloning and characterization of a mouse homologue (mNthl1) of  
 Escherichia coli endonuclease III.";  
 RL J. Mol. Biol. 282:761-774(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Luna L., Bjoras M., Rognes T., Hoff E., Seeberg E.;  
 RT "Complete genomic DNA sequence of the Mus musculus endonuclease III  
 homologue I gene (NTH1)."  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6J; TISSUE=Cecum;  
 RX MEDLINE-22354683; PubMed-12468851;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nakai I., Osato N., Saito R., Suzuki H., Yamana K.I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Buit C., Hume D.A., Quackenbush J.;

RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chochis C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Kongaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,  
RA Viterbo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilmink L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zengelen M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozone-Kishikawa T., Kono H., Nakamura M., Sakazume N., Seto K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa A.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RT Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RN Nucleotide 420:563-573(2002).  
RP SUBCELLULAR LOCATION.  
RX MEDLINE:22419222; PubMed:12531031.  
RA Ikeda S., Kohmoto T., Tabata R., Seki Y.;  
RT "Differential intracellular localization of the human and mouse  
RT endonuclease III homologs and analysis of the sorting signals.";  
CC -1- FUNCTION: Has both an apurinic and/or apyrimidinic endonuclease  
CC activity and a DNA N-glycosylase activity. Incises damaged DNA at  
CC cytosines, thymine and guanines. Acts on a damaged strand, 5'  
CC DNA damage and spontaneous mutagenic lesions.  
CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or  
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,  
CC leaving a 3'-terminal unsaturated sugar and a product with a  
CC terminal 5'-phosphate.  
CC -1- COFACTOR: Binds a 4Fe-4S cluster which is not important for the  
CC catalytic activity, but which is probably involved in the proper  
CC positioning of the enzyme along the DNA strand (By  
CC similarity).  
CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
CC -1- TISSUE SPECIFICITY: Widely expressed.  
CC -1- SIMILARITY: BELONGS TO THE NTH/MYH FAMILY.  
CC -1- CAUTION: Ref.3 sequence differs from that shown due to a  
CC frameshift in position 33.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AB006812; BAA22080.1; -  
DR EMBL: AB000371; BAA28846.1; -  
DR EMBL: AJ001617; CAB5239.1; -  
DR EMBL: Y09688; CAA70866.1; -  
DR EMBL: AK033701; BAC28435.1; ALT\_FRAME.  
DR MGI: MGI:131375; NCH11.  
DR GO: GO:0005739; C:mitochondrion; IDA.  
DR GO: GO:0005634; C:nucleus; IDA.  
DR GO: GO:0019104; F:DNA N-glycosylase activity; IDA.  
DR GO: GO:0006296; P:nucleotide-excision repair, DNA incision, 5. . .; IDA.  
DR InterPro: IPR003265; Endo3c.  
DR InterPro: IPR004036; EndoIII\_HhH.  
DR InterPro: IPR003651; FES\_Bind.  
DR InterPro: IPR004445; HhH.

DR Pfam: PF00730; HhH-GPD; 1.  
DR Pfam: PF00633; HhH; 1.  
DR SMART: SM00478; ENDO3c; 1.  
DR SMART: SM00525; FES; 1.  
DR PROSITE: PS00764; ENDO\_NUCLEASE\_III\_1; FALSE\_NEG.  
DR PROSITE: PS01155; ENDO\_NUCLEASE\_III\_2; 1.  
KW Hydroxylase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;  
KW Glycosylase; Lyase; Iron-sulfur; 4Fe-4S; Mitochondrion  
FT ACT\_SITE 208  
FT METAL 278  
FT METAL 285  
FT METAL 288  
FT METAL 294  
FT CONFLICT 112  
SQ SEQUENCE 300 AA; 33646 MW; A7D771FFD7BC1939 CRC64;  
Query Match 9.5%; Score 161.5; DB: 1; Length 300;  
Best Local Similarity 25.8%; Pred. No. 1.4e-05;  
Matches 56; Conservative 41; Mismatches 87; Indels 33; Gaps 6;  
QY 11 WYREARLPWRGKDP-----YRVLVSEVLLQOTRVEGALPY 49  
Db 80 WQOLANIRIMRSKRDAPVDQLGAENHCYDASAPPKVRVQVLSLMSQTKDOVTAGAM 139  
QY 50 RRELEPPYLKALAAASLEFVLRWMOAGYR-RAEHLRLARSE-----ELPPFAEL 103  
Db 140 QRLARCLFVESILQTDQDTGLRLTPVGFMRKVKYIKQTVALIQORVEGDIPASVAEL 199  
QY 104 RGLPGLPYTAANAASIAEGE-RVAAVDGNVRVLSRLFARESPEKEFLFALAGLPEG 162  
Db 200 VALPVGPKRAHLMAVAAMGTSIGIAVDTHVIRIANRL-RMTKMKRTTPEETKNIIEW 257  
QY 163 VDPGVW---NQALMELGATVCLPKRPGACPLGAF 196  
Db 258 LPRVLWSEVNGLLVGRGQGLCPVHPROACLNALC 294

Search completed: September 15, 2003, 13:51:14  
job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 13:49:19 : Search time 30 Seconds  
(without alignments)  
458.368 Million cell updates/sec

Title: US-09-938-901-2

Perfect score: 1700  
Sequence: 1 VEAWRKALLAWYRENNARPLP.....VLRKALPLLAHAGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/laa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/laa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/laa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/laa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/laa/PTGUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/laa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	505	29.7	403	4	US-09-107-532A-7308
2	484.5	28.5	535	3	US-08-813-574-2
3	477.5	28.1	470	4	US-09-252-991A-2310
4	477	28.1	375	4	US-09-198-452A-421
5	471.5	27.7	350	3	US-08-813-574-9
6	471.5	27.7	350	4	US-09-651-656-37
7	471.5	27.7	350	4	US-09-650-855-37
8	443	26.1	360	4	US-09-134-001C-4684
9	442	26.0	346	4	US-09-328-352-4982
10	274.5	16.1	221	1	US-08-663-023-17
11	274.5	16.1	221	4	US-09-402-959A-10
12	188	11.1	227	4	US-09-107-532A-3843
13	176	10.4	207	2	US-08-808-550-36
14	176	10.4	259	2	US-08-808-550-39
15	172	10.1	209	2	US-08-808-550-31
16	171	10.1	204	2	US-08-808-550-32
17	164.5	9.7	211	2	US-08-808-550-37
18	164.5	9.7	257	2	US-08-808-550-42
19	164.5	9.7	224	4	US-08-808-550-2
20	160	9.4	224	2	US-09-134-001C-3225
21	154	9.1	213	2	US-08-808-550-35
22	145	8.5	207	2	US-08-808-550-29
23	145	8.5	211	2	US-08-808-550-38
24	145	8.5	211	4	US-09-651-656-25
25	145	8.5	211	4	US-09-650-855-25
26	134	7.9	207	2	US-08-808-550-30
27	128	7.5	628	4	US-09-252-991A-30904

28	127.5	7.5	212	4	US-09-252-991A-17572
29	121	7.1	4472	2	US-08-804-227C-2
30	118.5	7.0	776	4	US-09-252-991A-28446
31	117	6.9	230	4	US-09-328-352-5770
32	115.5	6.8	494	4	US-09-252-991A-17295
33	115	6.8	542	4	US-09-252-991A-21793
34	115	6.8	702	4	US-09-252-991A-31609
35	114.5	6.7	425	4	US-09-252-991A-26326
36	113	6.6	316	4	US-09-252-991A-30325
37	111	6.5	675	4	US-09-252-991A-29650
38	111	6.5	1729	4	US-09-553-690-2
39	109.5	6.4	1580	2	US-08-804-227C-11
40	109.5	6.4	1580	2	US-08-804-198-5
41	108.5	6.4	460	4	US-09-328-352-4274
42	107.5	6.3	427	4	US-09-252-991A-30434
43	106	6.2	435	4	US-09-252-991A-24914
44	106	6.2	798	4	US-09-252-991A-25309
45	105.5	6.2	657	4	US-09-252-991A-28001

## ALIGNMENTS

RESULT 1  
US-09-107-532A-7308  
Sequence 7308, Application US/09107532A  
Patent No. 6583275

GENERAL INFORMATION:  
APPLICANT: Lynn A Doucelte-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FACIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS: ENTEROCOCCUS CORPORATION

ADDRESS: 100 Beaver Street  
City: Waltham  
State: Massachusetts  
Country: USA

ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107, 532A

PRIOR APPLICATION DATA:  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Arianello, Pamela Deneke  
REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-8277  
TELEFAX: (781)893-5007

INFORMATION FOR SEQ ID NO: 7308:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...403

SEQUENCE DESCRIPTION: SEQ ID NO: 7308:  
US-09-107-532A-7308

Sequence 17572, A  
Sequence 2, Appl1  
Sequence 28446, A  
Sequence 5770, Ap  
Sequence 17295, A  
Sequence 21793, A  
Sequence 31609, A  
Sequence 26326, A  
Sequence 30325, A  
Sequence 29650, A  
Sequence 2, Appl1  
Sequence 11, Appl1  
Sequence 5, Appl1  
Sequence 30434, Ap  
Sequence 24914, A  
Sequence 25309, A  
Sequence 28001, A

Query Match 29.7%: Score 505; DB 4; Length 403;  
 Best Local Similarity 32.7%: Pred. No. 1,4e-42;  
 Matches 118; Conservative 66; Mismatches 113; Indels 64; Gaps 11;

```

OY 9 LAMVRENARPLPMRGKEDPYRVLSEVLLQOTRVEQALPYRRRLERFPPLKLAASLE 68
DB 32 IOMYBEKRNLFWRNRPDIWIMSEIMLQOTRDYIDFYRMEFPIEELANAPPE 91
OY 69 EYLRAWGAGYRRAEHLRLARSV-----ELPPSAELRGLPGDGYTAAVAASIAFG 123
DB 92 KILKAMEGLGYSSRAARNIOAAKQIMSEFGEEMQPEEISLKGIGPYTGALASIAFG 151
OY 124 ERVAADVGNVRVLSRLFARES-----PKERELFALAOGLLPGGVDPGVNOMLGLATV 179
DB 152 LPEPAVDGNVRVLSRLFARES-----PKERELFALAOGLLPGGVDPGVNOMLGLATV 179
OY 180 CLKPRPGACPLGAFG-----RGKAPGRYPA-PRKRAKKEERLVALVLLGRKGVH---- 230
DB 212 CTPLSPKCECPLOAFCLANKRGIOF--SFVYTKAKPKOVYIYISALQNHSGAYYFEE 269
OY 231 -----LEREGFQGLGVPLF-----PPEELP-GREAFGV 261
DB 270 RDSQKLLANMWPPEMEVYQOEYERLKKEMETKEIDLDLVAEEONLPFEKQELFW 329
OY 262 RSRPGEVHALTHRLRLEV- RGALMEGEGED-----PKRRP-----LPKIMEKV 306
DB 330 OTHRLGEVTHIFSLKMHVLLFYGRATEGAQEETENKTSKMLPDAFDSVVPKQVKML 389
OY 307 L 307
DB 390 V 390

```

RESULT 2  
 US-08-813-574-2

```

; Sequence 2, Application US/08813574
; Patent No. 6013473
; GENERAL INFORMATION:
; APPLICANT: Wei, Yung-Fei
; TITLE OF INVENTION: Human Muty
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/813,574
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/013,132
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 535 amino acids
; TYPE: amino acid

```

STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-813-574-2

Query Match 28.5%: Score 484.5; DB 3; Length 535;  
 Best Local Similarity 36.8%: Pred. No. 2.5e-40;  
 Matches 136; Conservative 46; Mismatches 107; Indels 83; Gaps 13;

```

OY 1 VAMRKALLAMVRENARPLPMRG-----ENDPYRVLSVLLQOTRVEQALPYRRF 52
DB 79 VTAERGSLSWYDDEKRDLPNRRARDEDDDLRAVAWSEVMLOOTQVATVINTYTGW 138
OY 53 LERFPTLKAALAAASLEEVNLWOGAGYRRAEHLRLARS-VEEL---PSPFAL--RGL 106
DB 139 MQKWPLODLASLSLEEVNLWOGAGYRSGRRRLQEGARKKYVEELGGHMRTRAEYLQQL 198
OY 107 PGCLPYTAAVAASIAFERVAADVGNVRVLSRLFA-----RESPKEELFALAOGLPE 161
DB 199 PGVGRYTAGIASIAFGATGVGVNVARVLCRVRAIGADPSSITLVSOQLMGIAQOLV-D 257
OY 162 GVDPGVWNOALMELGATVCLPKPRPGACPLGAFGRKE----- 200
DB 258 PARPGDFNOAMELGATVCTPDRPLCSQCPVESICRARQVREQEQLLAGSGSSPDVEE 317
OY 201 -AP-----GRYPARRRRRAK---EERLVALV-----LGRKGVHLE 232
DB 318 CAPNTGQCHLCLPSEPMDQTLGVNFPKRASKRPRESSATCYLEPGALGAQILLVQ 377
OY 233 R-LEGRFGILYGP--LFPPEELPGRFAF-----GVNSRPLGEVHALTHRL 278
DB 378 RNSGLLAGLMEFSPVWSPERSLOLRKALLQELQWAGPLPATHLRLHGEVYHFSHIKL 437
OY 279 RVEVRGALMEGE 290
DB 438 TYGVYGLALEGQ 449

```

RESULT 3  
 US-09-252-991A-23310

```

; Sequence 23310, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23310
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-23310

```

Query Match 28.1%: Score 477.5; DB 4; Length 470;  
 Best Local Similarity 36.2%: Pred. No. 1.1e-39;  
 Matches 126; Conservative 58; Mismatches 119; Indels 45; Gaps 14;

```

OY 2 EAMRKALLAMVRENARPLPMRGKEDPYRVLSEVLLQOTRVEQALPYRRFLERFPPLK 60
DB 119 EGFNGAVLDMDYDRGRDLPMDQGITPRVWVSEIMLQOTQVSTYLGTFDFRMAALPDVE 178
OY 61 ALAASLEEVLRWOGAGYRRAEHLRLA-----RSVEELPSPFAELRGLPGDGYTA 115
DB 179 ALAAADEVLLHMTGLGYSRAARNLHKTQALVERHAGEPRPDVQLAEPLGIGSTAG 238
OY 116 AVASIAFERVAADVGNVRVLSRLFARES-----SPK-EKELFALAOGLLPGGVDPGVNMO 170

```

Db 239 AIALSLGKLRAPILDDGVKRVRLARYLAQDGPGBPKVRAALMEAEFTPH-ARVNHVYQ 297  
Qy 171 ALMELGATVCLPRPRGACPLGAFCR---GKEAPERYAPPKRRA-KEERLVALVLLG 225  
Db 298 AMMDLGTATLTGRKPSCLLPVSGCRHLLGRAD--YPOPKRKLPOKRLMPLIAN 355  
Qy 226 RKG---VHLERLGRFOGLYGVPLFPEELPGREAFGVRSPRLGEVR-----HALJTHR 276  
Db 356 RDAAILLYRRPSSGLMGLMSLPEL--DDLDGLEPLAARSLALGERRELSGLTHRTSHF 413  
Qy 277 RLVE-----VRGA-----LMEGEDPMKRPRLPKLMEVLRKA 310  
Db 414 QLAIEPLVAVGACAPRAVAGCDMLWYLAATPP-RLGLAAPVKKLKKA 460

RESULT 4  
US-09-198-452A-421  
; Sequence 421, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffiths, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 421  
; LENGTH: 375  
; TYPE: PRF  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-421

Query Match 28.1%; Score 477; DB 4; Length 375;  
Best Local Similarity 38.6%; Pred. No. 8,6e-40;  
Matches 117; Conservative 44; Mismatches 112; Indels 30; Gaps 7;

Qy 6 KALLARYENARLPWGEKDPYRVLYSEVLLQOTRVEQALPYRRLEPPTLKALAA 65  
Db 22 EALKWFEKKRSLPNDNPTYSVWYSEVLLQOTRAEVVIDYFNOMMEREPPTIESLAA 81  
Qy 66 SLEEVLRVMQAGYRRARHLRLARSVEE---LPSFAELRGJLGLGPTYAAVASI 120  
Db 82 KEDVDKLLWEGLLYSRAHLLGCAKRVMEEPHGKIPDDAISLAQIRGVGPTVHAILAF 141  
Qy 121 AFGERYAADVGNVRYLSRLFARESPEKE---LFALAQGLPEGVDPGVNQAALMEL 175  
Db 142 AFRRAAADVGNVRLVLSRIFLETISDLESTRTWVSRIQAALLPH-KSPRYLAELIEL 200  
Qy 176 GATVCLPRPRGACPLGAFCRGKEAPRGYRPAKRRRAKE---ERLVALVL-GRGVH 230  
Db 201 GATCIC-KYVQOCHRCVRCQACGAMREKQFVLPVRHAKKVFILHRLVAIVLDGSLVE 259  
Qy 231 LELLEGFOGLYGVPLFPEELPGREAFGVRSPRLGEVR-----PLGEVNHALLHR 277  
Db 260 KRRPKEMAGLYEFPYIEVEPEEGLODIEGFTKKMELSLSPLEIGNLKEORHAFTNHK 319  
Qy 278 LRV 280  
Db 320 VHL 322

RESULT 5  
US-08-813-574-9  
; Sequence 9, Application US/08813574  
; Patent No. 6013473  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Ying-Fel  
; TITLE OF INVENTION: Human Muty  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/813,574  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/013,132  
FILING DATE: 11-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Han, William T  
REGISTRATION NUMBER: 34,344  
REFERENCE/DOCKET NUMBER: ATG50002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5219  
TELEFAX: 610-270-4026  
TELEX:  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 350 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-813-574-9

Query Match 27.7%; Score 471.5; DB 3; Length 350;  
Best Local Similarity 35.1%; Pred. No. 2,8e-39;  
Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

Qy 8 LLARYENAR-PLWGEKDPYRVLYSEVLLQOTRVEQALPYRRLEPPTLKALAA 66  
Db 10 VLVDYRGKRLTPWQIDKTPYKWLSEVLLQOTQVATVPIYFERFARPTVVDLANAP 69  
Qy 67 LEEVLRVMQAGYRRARHLRLARSVEE---LPSFAELRGJLGLGPTYAAVASIA 121  
Db 70 LDEVLRHMTGIGYARRARHRLKAAQVATLHGKFPETFEVVALPVGSTAGAILSL 129  
Qy 122 AFGERYAADVGNVRYLSRLFARES-PEKE---LFALAQGLPE-GVDPGVNQAALMEL 175  
Db 130 LGRHFTPLDDGVKRVRLARYLAQDGPGBPKVRAALMEAEFTPH-ARVNHVYQ 297  
Qy 176 GATVCLPRPRGACPL-GAFCRGKEAPRGYRPAKRRRAKEERLVALVLLGRGVH-- 231  
Db 188 GAMICTSRKPKCSLCPLONCICAAANNSMALYPCKKPKQYLPERTGYFLLQHEDEVLLA 247  
Qy 232 -ERLEGFOGLYGVPLFPEPE---LPGREAFGVRSPRLGEVNHALLHR----- 280  
Db 248 QRPSPGLMGLYCEPQFADDESLRQWLAQRIADNLTQ-LTARHRTFSHFHLDIVPWL 306  
Qy 281 -----EVRLAMEGEDP---WKRPPLKMEVLRKALPL 313  
Db 307 PVSSFTGCMDEGNALWYNLAQPSVGLAAVYERLQO-LRTGAPV 350

RESULT 6  
US-09-651-656-37  
; Sequence 37, Application US/09651656  
; Patent No. 6340566  
; GENERAL INFORMATION:  
; APPLICANT: MCCUTHEN-MALONEY, SANDRA  
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY  
; TITLE OF INVENTION: DETECTION AND QUANTIFICATION OF SINGLE NUCLEOTIDE  
; TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,

;; TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES  
;; FILE REFERENCE: IL-10689

;; CURRENT APPLICATION NUMBER: US/09/651,656  
;; CURRENT FILING DATE: 2000-08-29

;; PRIOR APPLICATION NUMBER: 60/192,764  
;; PRIOR FILING DATE: 2000-03-28

;; NUMBER OF SEQ ID NOS: 106  
;; SOFTWARE: Patent In Ver. 2.1

;; SEQ ID NO: 37

;; LENGTH: 350

;; TYPE: PRN  
;; ORGANISM: Escherichia coli

US-09-651-656-37

Query Match  
Best Local Similarity 27.7%; Score 471.5; DB 4; Length 350;  
Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAAYRENAR-PLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66

DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVLMLOOTVATVYFERFMRPFVTVDLANAP 69

QY 67 LEEVLYRWOGAGYRRAEHLHRLARSVEEL-----PSPFAELRGLPGIGPTAAVASIA 121

DB 70 LDEVLYHMTGLGYAARNLHRAAOQVATLHGKRFETFEVVALPGVGRSTAGAILSTLS 129

QY 122 FGERVAADVGNVRVLSRLFARES--PKRE-----LFALAQGLLPE-GVDEGVNQALMEL 175

DB 130 LGHNFPLDGNVRYLARCAYASGMPKKEVENKIMLSLSEGVPAVER--FNQAMMDL 187

QY 176 GATVCLPKPRCGACPL--GAFCKGKEAPGRYPARRKRAKEERLVALYLGRKGYHL-- 231

DB 188 GAMICTRSKPKSLCPDLONGCIAANNMSMALYPGKKPKQTLPERTGYFLLQHEDEVLA 247

QY 232 -ERLEGRGGLGYVPLFPPEE-----LPGREAGVRSRLGVRNALTTHRLRV----- 280

DB 248 QRPSSGLMGGLYCFQFADDESLNQWLAQRIADNLTO-LTAFRHTFSHFHLDIVPMWL 306

QY 281 -----EVRGALMBEGEDP---WKRPPLKMEKVLKALPL 313

DB 307 PVSSFTGCMDEGNALMTNLNLAOPPSVGLAAPERLLOO-LRTGAPV 350

US-09-650-855-37

Query Match  
Best Local Similarity 35.1%; Pred. No. 2,8e-39;  
Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAAYRENAR-PLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66

DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVLMLOOTVATVYFERFMRPFVTVDLANAP 69

QY 67 LEEVLYRWOGAGYRRAEHLHRLARSVEEL-----PSPFAELRGLPGIGPTAAVASIA 121

DB 70 LDEVLYHMTGLGYAARNLHRAAOQVATLHGKRFETFEVVALPGVGRSTAGAILSTLS 129

QY 122 FGERVAADVGNVRVLSRLFARES--PKRE-----LFALAQGLLPE-GVDEGVNQALMEL 175

DB 130 LGHNFPLDGNVRYLARCAYASGMPKKEVENKIMLSLSEGVPAVER--FNQAMMDL 187

QY 176 GATVCLPKPRCGACPL--GAFCKGKEAPGRYPARRKRAKEERLVALYLGRKGYHL-- 231

DB 188 GAMICTRSKPKSLCPDLONGCIAANNMSMALYPGKKPKQTLPERTGYFLLQHEDEVLA 247

QY 232 -ERLEGRGGLGYVPLFPPEE-----LPGREAGVRSRLGVRNALTTHRLRV----- 280

DB 248 QRPSSGLMGGLYCFQFADDESLNQWLAQRIADNLTO-LTAFRHTFSHFHLDIVPMWL 306

QY 281 -----EVRGALMBEGEDP---WKRPPLKMEKVLKALPL 313

DB 307 PVSSFTGCMDEGNALMTNLNLAOPPSVGLAAPERLLOO-LRTGAPV 350

US-09-650-855-37

Query Match  
Best Local Similarity 35.1%; Pred. No. 2,8e-39;  
Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAAYRENAR-PLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66

DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVLMLOOTVATVYFERFMRPFVTVDLANAP 69

QY 67 LEEVLYRWOGAGYRRAEHLHRLARSVEEL-----PSPFAELRGLPGIGPTAAVASIA 121

DB 70 LDEVLYHMTGLGYAARNLHRAAOQVATLHGKRFETFEVVALPGVGRSTAGAILSTLS 129

QY 122 FGERVAADVGNVRVLSRLFARES--PKRE-----LFALAQGLLPE-GVDEGVNQALMEL 175

DB 130 LGHNFPLDGNVRYLARCAYASGMPKKEVENKIMLSLSEGVPAVER--FNQAMMDL 187

QY 176 GATVCLPKPRCGACPL--GAFCKGKEAPGRYPARRKRAKEERLVALYLGRKGYHL-- 231

DB 188 GAMICTRSKPKSLCPDLONGCIAANNMSMALYPGKKPKQTLPERTGYFLLQHEDEVLA 247

QY 67 LEEVLYRWOGAGYRRAEHLHRLARSVEEL-----PSPFAELRGLPGIGPTAAVASIA 121

DB 70 LDEVLYHMTGLGYAARNLHRAAOQVATLHGKRFETFEVVALPGVGRSTAGAILSTLS 129

QY 122 FGERVAADVGNVRVLSRLFARES--PKRE-----LFALAQGLLPE-GVDEGVNQALMEL 175

DB 130 LGHNFPLDGNVRYLARCAYASGMPKKEVENKIMLSLSEGVPAVER--FNQAMMDL 187

QY 176 GATVCLPKPRCGACPL--GAFCKGKEAPGRYPARRKRAKEERLVALYLGRKGYHL-- 231

DB 188 GAMICTRSKPKSLCPDLONGCIAANNMSMALYPGKKPKQTLPERTGYFLLQHEDEVLA 247

QY 232 -ERLEGRGGLGYVPLFPPEE-----LPGREAGVRSRLGVRNALTTHRLRV----- 280

DB 248 QRPSSGLMGGLYCFQFADDESLNQWLAQRIADNLTO-LTAFRHTFSHFHLDIVPMWL 306

QY 281 -----EVRGALMBEGEDP---WKRPPLKMEKVLKALPL 313

DB 307 PVSSFTGCMDEGNALMTNLNLAOPPSVGLAAPERLLOO-LRTGAPV 350

US-09-134-001C-4684

Query Match  
Best Local Similarity 34.3%; Pred. No. 2,2e-36;  
Matches 103; Conservative 56; Mismatches 119; Indels 22; Gaps 7;

QY 3 AMRKALLAWYRENAR-PLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKAL 62

DB 19 SFKKDIEDWPHKQDRMPHRETTNPPYIMLSEVLMLOOTVATVYFERFMRPFVTVDLANAP 69

QY 63 AASLEEVLYRWOGAGYRRAEHLHRLARSVE-----ELPSPFAELRGLPGIGPTAAVASIA 121

DB 79 SEANEDDEVLYHMTGLGYAARNLHRAAOQVATLHGKRFETFEVVALPGVGRSTAGAILSTLS 129

QY 118 ASIAFGERVAADVGNVRVLSRL--FARESPEKELFALAQGLLPE-GVDEGVNQALMEL 174

DB 139 MSAFENHPLATVDGNVRYLARCAYASGMPKKEVENKIMLSLSEGVPAVER--FNQAMMDL 187

QY 175 LGATVCLPKPRCGACPL--GAFCKGKEAPGRYPARRKRAKEERLVALYLGRKGYHL-- 231

DB 199 LGALVCTPKSPCLCPDLONGCIAANNMSMALYPGKKPKQTLPERTGYFLLQHEDEVLA 247

QY 232 -ERLEGRGGLGYVPLFPPEE-----LPGREAGVRSRLGVRNALTTHRLRV----- 280

DB 259 LKREKELKMGWQFPM--REQTNAVNDVSDDGSIETINERVFELKHQFTHLTWEIKV 316

US-09-328-352-4982

Query Match  
Best Local Similarity 34.3%; Pred. No. 2,2e-36;  
Matches 103; Conservative 56; Mismatches 119; Indels 22; Gaps 7;

QY 3 AMRKALLAWYRENAR-PLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKAL 62

DB 19 SFKKDIEDWPHKQDRMPHRETTNPPYIMLSEVLMLOOTVATVYFERFMRPFVTVDLANAP 69

QY 63 AASLEEVLYRWOGAGYRRAEHLHRLARSVE-----ELPSPFAELRGLPGIGPTAAVASIA 121

DB 79 SEANEDDEVLYHMTGLGYAARNLHRAAOQVATLHGKRFETFEVVALPGVGRSTAGAILSTLS 129

QY 118 ASIAFGERVAADVGNVRVLSRL--FARESPEKELFALAQGLLPE-GVDEGVNQALMEL 174

DB 139 MSAFENHPLATVDGNVRYLARCAYASGMPKKEVENKIMLSLSEGVPAVER--FNQAMMDL 187

QY 175 LGATVCLPKPRCGACPL--GAFCKGKEAPGRYPARRKRAKEERLVALYLGRKGYHL-- 231

DB 199 LGALVCTPKSPCLCPDLONGCIAANNMSMALYPGKKPKQTLPERTGYFLLQHEDEVLA 247

QY 232 -ERLEGRGGLGYVPLFPPEE-----LPGREAGVRSRLGVRNALTTHRLRV----- 280

DB 259 LKREKELKMGWQFPM--REQTNAVNDVSDDGSIETINERVFELKHQFTHLTWEIKV 316

US-09-328-352-4982

Query Match  
Best Local Similarity 34.3%; Pred. No. 2,2e-36;  
Matches 103; Conservative 56; Mismatches 119; Indels 22; Gaps 7;

QY 3 AMRKALLAWYRENAR-PLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKAL 62

DB 19 SFKKDIEDWPHKQDRMPHRETTNPPYIMLSEVLMLOOTVATVYFERFMRPFVTVDLANAP 69

QY 63 AASLEEVLYRWOGAGYRRAEHLHRLARSVE-----ELPSPFAELRGLPGIGPTAAVASIA 121

DB 79 SEANEDDEVLYHMTGLGYAARNLHRAAOQVATLHGKRFETFEVVALPGVGRSTAGAILSTLS 129

QY 118 ASIAFGERVAADVGNVRVLSRL--FARESPEKELFALAQGLLPE-GVDEGVNQALMEL 174

DB 139 MSAFENHPLATVDGNVRYLARCAYASGMPKKEVENKIMLSLSEGVPAVER--FNQAMMDL 187

QY 175 LGATVCLPKPRCGACPL--GAFCKGKEAPGRYPARRKRAKEERLVALYLGRKGYHL-- 231

DB 199 LGALVCTPKSPCLCPDLONGCIAANNMSMALYPGKKPKQTLPERTGYFLLQHEDEVLA 247

QY 232 -ERLEGRGGLGYVPLFPPEE-----LPGREAGVRSRLGVRNALTTHRLRV----- 280

GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
TITLE OF INVENTION: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328.352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 4982  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-4982

Query Match 26.0%; Score 442; DB 4; Length 346;  
Best Local Similarity 33.3%; Pred. No. 2,6e-36;  
Matches 113; Conservative 60; Mismatches 118; Indels 48; Gaps 11;

QY 7 ALLAWYREMAR-PLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPPLKALAA 65  
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 11 ALLWYFOHGRHNDLPWYADDPYKVMYSEIMLOOTQVTVLQYEDRHEPPEYVALGYA 70  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 66 SLEEVLRWOGAGYVYRRAHLRLARSVE--LPPSFAELRGLPGIGPYTAAVAASIAF 122  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 71 TWDEVAPYMACLGYYARARNLHKAAGLVAOQKFPETLEEMIALPGIGRSTAGALMSGL 130  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 123 GERAAVDGNVRRVLSLFARES-----PKKEKELFALAQGLPEGVDGVMNOALMEGA 177  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 131 ROYVINDGNVRRVRLARFALIEDLSKPOHEREMWKLAEELCPTRHND-YTQALMDGA 189  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 178 TVCLPRKRCGACPLGAFAC-----RGKEAPGRYPAPRRKRAKEERLVALVLGKGVHLER 233  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 190 TICPRKPKCLICVMOAHCAVQOGLDELFPKPKPKPKTKTADVLIIQCEDEWFMOQR 249  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 234 -LEGREFGLGVPLPEPPE-ELPGREAAFGVRSRP-LGEVRHALTH----- 275  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 250 QAHGLMGGLFCLPILENEHERLKLQOQFKLPQPOQTFQISHSFHTFWLNAHVHVPD 309  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 276 --RLRVEVRGALMBEGBEDPWKRRLPKLMKVLKALP 312  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 310 QKEHLAIELEG-----QWLSP-----EQALAKGP 334  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |

RESULT 10  
US-08-663-023-17  
Sequence 17, Application US/08663023  
Patent No. 5763178  
GENERAL INFORMATION:  
APPLICANT: CHIRIKIJAN, Jack G.  
APPLICANT: COLLIER, G. Bruce  
TITLE OF INVENTION: OSCILLATING SIGNAL AMPLIFIER FOR NUCLEIC  
TITLE OF INVENTION: ACID DETECTION  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/663.023  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012.950  
FILING DATE: 06-MAR-1996  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/483.089  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 66669/110  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-663-023-17

Query Match 16.1%; Score 274.5; DB 1; Length 221;  
Best Local Similarity 31.7%; Pred. No. 1e-19;  
Matches 63; Conservative 40; Mismatches 85; Indels 11; Gaps 4;

QY 8 LLAWYREMARPLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPPLKALAA 67  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 16 ILTFWNTDRDPFPMHRTDPPVILITELIRRTTAGHVKKIYDKFPVKYKCFEDILKTPK 75  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 68 EEVLRWOGAGY-YRRAHLRLARSV-----ELPPSFAELRGLPGIGPTTAAVAASIA 121  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 76 SEIANDIEIGISNORAEQLKELARVINDYGGVRRNRKAILDLPGVKTYTCAAVMCLA 135  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 122 FGERAAVDGNVRRVLSLFARE-----SPKEKELFALAQGLPEGVDGVMNOALMEGA 177  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 136 FGKKAAMVDANFVRININIFGGSYENLNINIKALMELAEITLVPGKCHD-FRLGLMDISA 194  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 178 TVCLPRKRCGACPLGAFAC 196  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 195 TICAPRKPKCEKCKMSKLC 213  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |

RESULT 11  
US-09-402-959A-10  
Sequence 10, Application US/09402959A  
Patent No. 6548247  
GENERAL INFORMATION:  
APPLICANT: CHIRIKIJAN, Jack G.  
APPLICANT: BAZAR, LEONARD S.  
TITLE OF INVENTION: DETECTION AND MAPPING OF POINT MUTATIONS USING PARTIAL  
FILE REFERENCE: 066669/0118  
CURRENT APPLICATION NUMBER: US/09/402.959A  
CURRENT FILING DATE: 1999-12-09  
PRIOR APPLICATION NUMBER: PCT/US98/06878  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/043,184  
PRIOR FILING DATE: 1997-04-16  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 10  
LENGTH: 221  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: ORF 10  
US-09-402-959A-10

Query Match 16.1%; Score 274.5; DB 4; Length 221;  
Best Local Similarity 31.7%; Pred. No. 1e-19;  
Matches 63; Conservative 40; Mismatches 85; Indels 11; Gaps 4;

QY 8 LLAWYREMARPLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPPLKALAA 67  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 16 ILTFWNTDRDPFPMHRTDPPVILITELIRRTTAGHVKKIYDKFFVYKCFEDILKTPK 75  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |

QY 68 EVLVWOGAGY-YRAEHLRLARSV-----ELPPSFAELRGPLGPTAAVAASIA 121  
 Db 76 SEIANDIKEIGLSNRAPQKELRLARVINDYGGVFPNNRKALIDLEPGVGYTCAAAVWCLA 135  
 QY 122 FGERAAVADGNVNRVLSRLPARE-----SPKEKELFALAGLPEGVDPGVWQALMELCA 177  
 Db 136 FGKAAVADVNFVYVINKYFGGSYENLNHYKALMELAEITLVGKCRD-FNLGLMDFSA 194  
 QY 178 TVCLPKRRRCGACPLGAFNC 196  
 Db 195 IICAPRRKCEKCGMSKLC 213

## RESULT 12

US-09-107-532A-3843  
 ; Sequence 3843, Application US/09107532A  
 ; Patent No. 6583275  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 7310  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 ; STREET: 100 Beaver Street  
 ; CITY: Waltham  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02354  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-ROM ISO9660  
 ; COMPUTER: PC  
 ; OPERATING SYSTEM: <Unknown>  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/107,532A  
 ; FILING DATE: 30-Jun-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/085,598  
 ; FILING DATE: 14 May 1998  
 ; APPLICATION NUMBER: 60/051571  
 ; FILING DATE: July 2, 1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Aridello, Pamela Deneke  
 ; REGISTRATION NUMBER: 40,489  
 ; REFERENCE/DOCKET NUMBER: GTC-012  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (781)893-5007  
 ; TELEFAX: (781)893-8277  
 ; INFORMATION FOR SEQ ID NO: 3843:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 227 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHEICAL: YES  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Enterococcus faecium  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (8) LOCATION 1...227  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 3843:  
 US-09-107-532A-3843

Query Match 11.1%; Score 188; DB 4; Length 227;  
 Best Local Similarity 30.6%; Pred. No. 5.3e-11;  
 Matches 63; Conservative 36; Mismatches 79; Indels 28; Gaps 9;

QY 26 DPNRVLSVEVLQ---TRVEQALPYRRFLRPFLTKALAASLEEVLRWOGAGYR- 81  
 Db 31 NPFELLLAVILSAQATDVSVNKATP---DLFASPTPDALAEASIDELIILKITITGLYRN 87  
 QY 82 RAEHLRLARSVE-----LPPSFAELRGPLGPTAAVAASIAFGERVAADVGNVRRV 136

Db 88 KAKINAKAQOOLIERFDGQVPTSRBELMSLPVGRKTANVVGDAFGIPAIAVDTHVERV 147  
 QY 137 LSRLEFARSPKEKELFALAGLPEGVDPGVW---NOALMELGATVCLPKRRRCACPLG 193  
 Db 148 SKRL--RICKLDATVMEVEETLMRK-VPELWVKTHHTLIFGRYHCTARNPKCEVPL 204  
 QY 194 AFGR-GKEAPGRYPAPRRKAREERL 218  
 Db 205 SICODGK-----NRMLKEKAL 221

## RESULT 13

US-08-808-550-36  
 ; Sequence 36, Application US/08808550  
 ; Patent No. 5871992  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Teebor, George W.  
 ; APPLICANT: Hilbert, Timothy P.  
 ; TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: David A. Jackson, Esq.  
 ; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
 ; CITY: Hackensack  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07601  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentln Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/808,550  
 ; FILING DATE: 26-FEB-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jackson Esq., David A.  
 ; REGISTRATION NUMBER: 26,742  
 ; REFERENCE/DOCKET NUMBER: 1049-1-001 N  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201-487-5800  
 ; TELEFAX: 201-343-1684  
 ; INFORMATION FOR SEQ ID NO: 36:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 207 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHEICAL: NO  
 ; FRAGMENT TYPE: N-terminal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Escherichia coli  
 US-08-808-550-36

Query Match 10.4%; Score 176; DB 2; Length 207;  
 Best Local Similarity 32.0%; Pred. No. 7.6e-10;  
 Matches 58; Conservative 33; Mismatches 70; Indels 20; Gaps 8;

QY 28 YVIVLSEVLQDTRVEQALPYRRFLRPFLTKALAASLEEVLRWOGAGYR-RAEHL 86  
 Db 29 FOVLVAIWLSSQTRDEVAAAMKRLKDHGSLGKILIEFKVPDLEITLCLCVGFKKRAVYL 88  
 QY 87 HRLARSVE-----ELPPSFAELRGPLGPTAAVAASIAFGERVA-ADVGNVRRVLSRL 140  
 Db 89 OKTAKIILKDDSGGIPISLGLCALPEVGPKNANLWQJAMGECVGIADVTHHRSNRL 148  
 QY 141 -FARSPKEKELFALAGLPEGVDPGVW---NOALMELGATVCLPKRRRCACPLGAFNC 196





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 13:48:09 ; Search time 40 Seconds

(without alignments)  
781.371 Million cell updates/sec

Title: US-09-938-901-2

Perfect score: 1700

Sequence: 1 VEAMRKALLAWYRENARPLP.....VLKRALPLLAAGVPLPDA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR-76:\*

2: PIR1:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	669.5	39.4	363	2	A/G-specific adeni
2	551.5	32.4	375	2	A/G-specific adeni
3	520	30.6	349	2	A/G-specific adeni
4	515.5	30.3	367	2	A/G-specific adeni
5	515.5	30.3	367	2	A/G-specific adeni
6	503	29.6	349	2	A/G-specific adeni
7	501.5	29.5	369	2	A/G-specific adeni
8	498	29.3	349	2	A/G-specific adeni
9	497.5	29.3	372	2	A/G-specific adeni
10	492	28.9	387	2	A/G-specific adeni
11	491.5	28.9	350	2	A/G-specific adeni
12	491	28.9	371	2	A/G-specific adeni
13	484.5	28.5	350	2	A/G-specific adeni
14	483	28.4	365	2	A/G-specific adeni
15	478.5	28.1	353	2	A/G-specific adeni
16	478	28.1	365	2	A/G-specific adeni
17	477.5	28.1	355	2	A/G-specific adeni
18	477	28.1	369	2	A/G-specific adeni
19	477	28.1	369	2	A/G-specific adeni
20	476	28.0	339	2	A/G-specific adeni
21	471.5	27.7	350	2	A/G-specific adeni
22	469.5	27.6	350	2	A/G-specific adeni
23	469.5	27.6	350	2	A/G-specific adeni
24	467	27.5	345	2	A/G-specific adeni
25	459.5	27.0	381	2	A/G-specific adeni
26	459.5	27.0	391	2	A/G-specific adeni
27	458.5	27.0	369	2	A/G-specific adeni
28	445	26.2	378	2	A/G-specific adeni
29	442.5	26.0	461	2	A/G-specific adeni

30	442.5	26.0	461	2	A/G-specific adeni
31	441.5	26.0	371	2	A/G-specific adeni
32	436	25.6	357	2	A/G-specific adeni
33	423	24.9	328	2	A/G-specific adeni
34	420	24.7	328	2	A/G-specific adeni
35	414	24.4	312	2	A/G-specific adeni
36	379	22.3	308	2	A/G-specific adeni
37	378.5	22.3	350	2	A/G-specific adeni
38	374.5	22.0	608	2	A/G-specific adeni
39	365.5	21.5	304	2	A/G-specific adeni
40	360	21.2	297	2	A/G-specific adeni
41	342	20.1	277	2	A/G-specific adeni
42	313.5	18.4	286	2	A/G-specific adeni
43	305.5	18.0	215	2	A/G-specific adeni
44	274.5	16.1	221	2	A/G-specific adeni
45	227	13.4	213	2	A/G-specific adeni

ALIGNMENTS

RESULT 1

A/G-specific adenine glycosylase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 08-Dec-2000

C:Accession: A75294

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

M.; Shen, M.; Yamatehavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.J.

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: A75294

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-363 <WHI>

A:Cross-references: GB:AE002060; GB:AE00513; NID:96460082; PID:AAF11831.1; PID:9646

A:Experimental source: strain R1

A:Gene: DR2285

A:Map position: 1

C:Superfamily: A/G-specific adenine glycosylase

Query Match	39.4%;	Score 669.5;	DB 2;	Length 363;
Best Local Similarity	46.2%;	Pred. No. 9.7e-45;		
Matches 154;	Conservative 48;	Mismatches 96;	Indels 35;	Gaps 11;
Qy	1	VEAMRKALLAWYRENARPLPWR----	GEKDYRYLVSEVLLQOTRVQDALPYRRFLERF	56
Db	17	VGALRDLGLGWRDRAGRDLPWRUGDEGRDPYRWVAETLLQOTQVARGSLERYERLEAF	76	
Qy	57	PTLKALAAALEEVLRWQAGYTYRAEHLRLARSVEE--LPPSFALRGLPLGPTYTA	114	
Db	77	PTYOALAAAPQDANLTKMBEGCGYARARNLHRAAIIIDEGFPDQYAGWMLALPEVGPTYTA	136	
Qy	115	AAVASIAFGRVAADGNVRVLSLRFARSPKKELEFALAGLLPESVDGPNQALME	174	
Db	137	AAVSSIALGEPRAVNDGNVRVLSLRLFAEHPSPDKWQEQDADRL--DPARPGANNEAVMD	195	
Qy	175	LGATVCLPKRRCACPLGAFRCRKE--AGRYAPARR--RAKERLVALVLGRKGVHL	231	
Db	196	LGATICVPSKSPACDRCPVSAHCAVYQGPDPAPAPARQARVRAVALLIGAEYAVL	255	
Qy	232	ERLEGR--FOGLYGVLPPEELPGREAAFGVRSR-----PLGEVHAULTHRLR	279	
Db	256	EKRSGSLGLFLGRLP---EIGARETAADALARLQARLGAEVKECCIGTQHGHTHRLS	312	
Qy	280	VEVRGALMEGEDPMPKR-----LPLKMEVL	307	
Db	313	VEV-----YRAEDRP--RQPVGAALSLRDKAL	340	





```
QY 1 VEAMRKALLAMYRENAARPLPWRGKEDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLK 60
D 13 IQQFRDLDLSMFREQGVLPWRDODPYKVMSEVMLQOTRVEVLPFLRFEPQEPYVE 72
QY 61 ALAAASLEEVLRVMOGAGYRRAEHLRLARSVE-----LPPSFAELRGPLGPGYTA 115
D 73 ALADADEEKLKMEGEGYYSRRNMQSAVKEVQEGYVPPDEKDFGGLKGVCPYTKG 132
QY 116 AVASIAFGERVAADVGNVRVLSRLFA---RESPKEKELFALAQGLLEPGVPGWNOA 171
D 133 AVLSIAVNNKRIPAVDGNVMVMYKRISINDDIKPPTRITFEDAIATAFISKEKPEFNOG 192
QY 172 LMEIGATVCLPKPRRCGACPLGAFCRGKEAPGRYPAPRRKRAKE---ERLVALVLLGRKG 228
D 193 LMEIGALICPKSPSCILCPVQOHCSAFEGTERELPVKSKKKPGIKTMAALVLDDEG 252
QY 229 ---VHLERLRQGLGVPLFPPEELPGREAFGR-----SRP 265
D 253 QVYIHKRPKSKGLLANLM-----FPPMLETQKGIKTEREQLIAELENEYGIQADISDL 304
QY 266 LGEVRHALTH 275
D 305 QGVVEHVETH 314
```

## RESULT 8

```
D81088
A/G-specific adenine glycosylase NMB1396 [Imported] - Neisseria meningitidis (strain MCS
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: D81088
R:Retcelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
rt, H.; Qin, H.; Yamathavan, J.; Gill, J.; Scarlato, V.; Masiagnan, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappunoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MCS8.
A:Reference number: AB10000; MUID:20175755; PMID:10710307
A:Accession: D81088
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <STO>
A:Cross-references: GB:AE002488; GB:AE002098; NID:g7226631; PIDN:AAFA1760.1; PID:g722663
A:Experimental source: serogroup B, strain MCS8
C:Genetics:
A:Gene: NMB1396
C:Superfamily: A/G-specific adenine glycosylase
```

```
Query Match 29.3%; Score 498; DB 2; Length 349;
Best Local Similarity 39.4%; Pred. No. 2,2e-31;
Matches 130; Conservative 52; Mismatches 112; Indels 36; Gaps 14;
```

```
QY 8 LLAWYRENAAR-PLPWRGKEDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66
D 14 LIRMQKQKGRHHLPMQ-VKNPYCWLSEIMLQOTRVAIVLOYIPRFLERFPTVQTLAAR 72
QY 67 LEEVLRVMOGAGYRRAEHLRLARSVE-----LPPSFAELRGPLGPGYTA 121
D 73 QDEVLSLWAGLGYSRRNLHKAQOVVROFGTFPSEKRLDLETLCGVGRSTAAICAFS 132
QY 122 FGERVAADVGNVRVLSRLFARE-SPKEKE---LEFALAQGLLEPGVPGWNOALMEL 175
D 133 FNRRTIILIDGNVKRYLCAVFAADGNPQKKENSLMTAESLLESENMDMPAYTQGLMDL 192
QY 176 GATVCLPKPRRCGACPLGAFCRGKEAPGRYPAPRRKRAKE---RLVALVLLGRKG-VHL 231
D 193 GATVCKRTKPLCHOCPMADICEAKQKNTAELPRKRTAAEVPJLPLVLYLRNNDGATLL 252
QY 232 ER--LEGRFOGLYGVPLFPPEELPGRE---AAGVRSRPLGE---VRHALTHRLRLEVVR 283
D 253 EKRPAGKIGWGLYCVPCF--ESLNGLSDFAAKFSITMADMDEQATLTHRLTHRLLLTTPF 310
```

```
QY 284 GALW--EGEGEDPWKR-----LPKIME 304
D 311 EAQMPSESPSDGIWKIPAHLDKDYGLPKPLE 340
```

## RESULT 9

```
C83766
adenine glycosylase BH0931 [Imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C83766
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: AB3650; MUID:20512582; PMID:11058132
A:Accession: C83766
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-372 <STO>
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04650.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0931
C:Superfamily: A/G-specific adenine glycosylase
```

```
Query Match 29.3%; Score 497.5; DB 2; Length 372;
Best Local Similarity 36.9%; Pred. No. 2,6e-31;
Matches 122; Conservative 54; Mismatches 120; Indels 35; Gaps 10;
```

```
QY 1 VEAMRKALLAMYRENAARPLPWRGKEDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLK 60
D 10 ISTFQNDLVTFWFSNHYRELPMRENKDPYRVVSEIMLQOTRVDIVIPYQAFMQPFTLE 69
QY 61 ALAAASLEEVLRVMOGAGYRRAEHLRLARSV-----ELPPSFAELRGPLGPGYTA 115
D 70 TLTAIEEOQVLAWEGLGYYSRRNLHKAQOVVROFGTFPSEKRLDLETLCGVGRSTAAICAFS 129
QY 116 AVASIAFGERVAADVGNVRVLSRLFARESPEKKELEFALAQGLLEPGV---DPGVNQA 171
D 130 AILSIAYDQPEPAVDGNVMVLSRVLYIEEDIAKVKTRITFESLTYLIDSKENSPFNOG 169
QY 172 LMEIGATVCLPKPRRCGACPLGAFCRGKEAPGRYPAD---RKRRAKEERLVALVLLGRKG 228
D 190 LMEIGALVCLPTSPGCLCPVRDHCRAFAAGVQDOLPIKAKKKKPAKQILAAVIRNEKG 249
QY 229 -VHLERL--EGRFOGLYGVPLFPPEELPGREAFGR---VRSR-----PLGE---VRHALT 274
D 250 QVLEIRPREKGLLAKMQFPVWELESTRKNAQOVVIGDYIHERFHLDAAVGEYVQTEHVFS 309
QY 275 HRLRLEVVRGALWEGEGEDPWKRPLPKLMEK 305
D 310 HLINWIRVYEATVKG-----VPSLNDK 331
```

## RESULT 10

```
A86730
purine nucleosidase (EC 3.2.2.1) muty [Imported] - Lactococcus lactis subsp. lactis (
N:Alternate names: A/G-specific adenine glycosylase
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: A86730
R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID:21225166; PMID:11337471
A:Accession: A86730
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-387 <STO>
A:Cross-references: GB:AE005176; PID:g12723767; PIDN:AAK04939.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: muty
```

[illegible]

A1:residues: 1-350 (UBS)  
A1:Cross-references: GB:M86634; NID:g295203; PIDN:AAA27165.1; PID:g154184

C:Genetics:  
A:Gene: mutB  
C:Superfamily: A/G-specific adenine glycosylase  
C:Keywords: DNA repair; iron-sulfur protein

[illegible]

```

RESULT 14
AD1657
A/G-specific adenine glycosylase homolog [ln1797] [imported] - Listeria innocua (strain
AD1657
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AD1657
R:Glaeson, P.; Fraenkel, I.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, T.; Dussutlet, O.; Entlan, K.D.; Eschl, H
Science 294, 849-852, 2001
A:Authors: Kreif, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; M
OK, C.; Schluerter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlard
A>Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1657
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-365 <GL>
A:Cross-references: GB:AL592022; PION:CMC97028.1; PID:g16414300; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: ln1797
C:Superfamily: A/G-specific adenine glycosylase
Query Match      28.4%; Score 483; De 2; Length 365;
Best Local Similarity 31.0%; Pred. No. 3.4e-30;
Matches 107; Conservative 65; Mismatches 121; Indels 52; Gaps 10;

QY      1 VEAMRKALLAWYRENAAPLPNGEKEDPYRVLSVYLLOQRVSOALPYVRRLERPTLK 60
       :|::||::||| | ||| :|||:||||::|||::|::|||::|::|||:
Db      14 IFAFGALASWYEANKRIILPMRENTEPEPIRWSEIIMLOQKVDTVPYFNRFTOPTME 73
QY      61 ALAAASLEVLAVNMGAGYYRAEHLHRLARSV-----ELPSFALRLGTLPGTGYTA 115
       :|::|||::||| | | | :| :| :| :| :| :| :| :| :| :| :| :|
Db      74 SFVADEADILKAMEDELGYTSVRNLQIRAMQWMTDFEGVEPSDLTTITLKGCGPRTAG 133
QY      116 AVASIFGEGRVAADVNCNRRVLSRLCFARF----SPKEKEFLALAQGLLPGCVDPGWNOA 171
       |:|::|||: |::|||::|||: |::|||: |::|||: |::|||: |::|||:
Db      134 AIISIAVANOAEPAVDGNNVMKVIARVLTSEIDIMAKSTKRIFEVLVYQLIDENPAAFNQG 193

```

```

RESULT 15
D82320
A/G-specific adenine glycosylase VC0452 [imported] - Vibrio cholerae (strain N16961 s
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82320
R:Heldberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.
Chadson, D.; Ermolaeva, M.D.; Yamathavan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483. 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833; PMID:10952301
A:Accession: D82320
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-353 <HIT>
A:Cross-references: GB:AE004132; GB:AE003852; NID:99654871; PIDN:AAE93625.1; GSPDB:CN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
C:Gene: VC0452
A:Map position: 1
C:Superfamily: A/G-specific adenine glycosylase

```

Query Match	28.1%	Score 478.5	DB 2	Length 353
Best Local Similarity	37.7%	Pred. No. 7.3e-30		
Matches 113	Conservative 49	Mismatches 109	Indels 29	Gaps 9
6 KALLAMYENARP-LPMRGCKDPRVLSEVLTLOOTRQDALPYRRFLERFPTLLAA	64			
6 QALITMYAYAGKNNLPMQONKNATRWLSSIMDQOVATVLYFERELFFPTVHALA	65			
65 ASLEEVLTVMQAGAGYRRRAELHRLARLSV-----ELPSPFAELRGLEGLPTAAVAS	119			
66 APODEVLHMTGLGYARARNLHKAOMVSEYGEFPTDLEQNNALRGVSTAAVL	125			
120 IAEGRVAAVDGNRRVLSLFLARE-----SPREKELFALAGLLPEGVDDGVNNQALME	174			
126 SYVKKPRHALIDGVKTKTLRCFAVDSGMGQKSVENQIMHNAEMHNRK-VYVDKTYQAMND	184			
175 LGATVCLPFRPPRCGACPLGAFCCRGKEA--PCRYAPARKRAKEERLVALVLGR-KGYHL	233			
185 MGAHICIRSKPSCSLCPRESFCLAQOQNPDEYTGKKPKTKDKVUKATWTVMLYHDAVWL	244			
232 ER--LEGRFOGLGVLPPEELPGREAAFGVRS-----RPLGEVNHALLTRKRLV	281			
245 EQRPGSGIMGGLT---CFQSGSEIANQOTIDORAIGDSTTSOKTTLAERHNTFSYHNIDI	30			

Search completed: September 15, 2003, 13:53:45  
Job time : 41 secs

Search completed: September 15, 2003, 13:53:45  
Job time : 41 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 19, 2005, 23:55:40 : Search time 111 Seconds  
(without alignments)  
6243.416 Million cell updates/sec

Title: US-09-938-901A-1  
Perfect score: 975  
Sequence: 1 atggagcgctgcgggaagc.....tagtccctcccgagcga 975

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/PCUTS\_COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	125.4	12.9	963	4	US-09-252-991A-6832 Sequence 6832, Ap
2	125.4	12.9	1401	4	US-09-252-991A-6886 Sequence 6886, Ap
3	125.4	12.9	1413	4	US-09-252-991A-6739 Sequence 6739, Ap
4	113	11.6	1551	4	US-09-489-039A-4788 Sequence 4788, Ap
5	107.2	11.0	549	4	US-09-252-991A-6783 Sequence 6783, Ap
6	104	10.7	1858	3	US-08-813-574-1 Sequence 1, Appl
7	87.2	8.9	2293	3	US-09-651-656-38 Sequence 38, Appl
8	87.2	8.9	2293	3	US-09-650-855-38 Sequence 38, Appl
9	86.6	8.9	4403765	3	US-09-103-840A-2 Sequence 2, Appl
10	86.6	8.9	4411529	3	US-09-103-840A-1 Sequence 1, Appl
11	82	8.4	1176	4	US-09-583-110-1922 Sequence 1922, Ap
12	82	8.4	9909	4	US-08-961-527-12 Sequence 12, Appl
13	81.2	8.3	968	4	US-09-864-866-46 Sequence 46, Appl
14	63.6	6.5	1212	4	US-09-107-532A-3654 Sequence 3654, Ap
15	62.4	6.4	1188	4	US-09-134-000C-1710 Sequence 1710, Ap
16	62.2	6.4	1227	3	US-09-385-028-23 Sequence 23, Appl
17	62.2	6.4	1227	4	US-09-726-614-23 Sequence 23, Appl
18	62.2	6.4	1227	4	US-09-385-028-23 Sequence 23, Appl
19	62.2	6.4	15079	3	US-09-385-040-23 Sequence 1, Appl
20	62.2	6.4	15079	4	US-09-726-614-1 Sequence 1, Appl
21	62.2	6.4	15120	4	US-09-385-040-1 Sequence 1, Appl
22	61.8	6.3	10444	4	US-09-543-681A-1143 Sequence 1143, Appl
23	61.4	6.3	1230025	4	US-09-198-452A-1 Sequence 1, Appl
24	59.6	6.1	20235	1	US-07-642-734C-3 Sequence 3, Appl
25	59.6	6.1	20235	3	US-08-439-009A-3 Sequence 3, Appl
26	58.8	6.0	1041	4	US-09-328-352-856 Sequence 856, Appl
27	56.6	5.8	2889	1	US-08-537-002A-4 Sequence 4, Appl

28	56.6	5.8	2889	3	US-08-863-010-4 Sequence 4, Appl
29	56.6	5.8	2889	3	US-09-024-429-4 Sequence 4, Appl
30	56.6	5.8	3600	1	US-08-537-002A-5 Sequence 5, Appl
31	56.6	5.8	3600	3	US-08-863-010-5 Sequence 5, Appl
32	56.6	5.8	3600	3	US-09-024-429-5 Sequence 5, Appl
33	55.6	5.7	2681	3	US-08-928-213B-7 Sequence 7, Appl
34	55.6	5.7	2363	4	US-09-818-780-22 Sequence 22, Appl
35	55.6	5.7	53500	4	US-09-266-965-76 Sequence 76, Appl
36	55.2	5.7	2634	3	US-08-941-936-1 Sequence 1, Appl
37	54.8	5.6	1404	4	US-09-252-991A-1725 Sequence 1725, Ap
38	54.8	5.6	2208	4	US-09-252-991A-2086 Sequence 2086, Ap
39	54.8	5.6	2586	4	US-09-252-991A-1804 Sequence 1804, Ap
40	54.8	5.6	3378	4	US-09-252-991A-2158 Sequence 2158, Ap
41	53.8	5.5	7560	4	US-09-754-112A-2 Sequence 2, Appl
42	53.8	5.5	12738	4	US-09-754-112A-1 Sequence 1, Appl
43	53.8	5.5	47981	4	US-09-679-279-1 Sequence 1, Appl
44	52.8	5.4	4257	2	US-08-690-473-1 Sequence 1, Appl
45	52.8	5.4	4257	3	US-09-259-821A-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-252-991A-6832  
Sequence 6832, Application US/09252991A  
Patent No. 6531795  
GENERAL INFORMATION:  
APPLICANT: MARC J. RUBENFELD et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252, 991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074, 788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094, 190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6832  
LENGTH: 963  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6832

Query Match 12.9%; Score 125.4; DB 4; Length 963;  
Best Local Similarity 54.7%; Pred. No. 2.5e-17;  
Matches 311; Conservative 0; Mismatches 231; Indels 27; Gaps 2;

QY	55	CTCCCTGCGCGGAGAGAACCTTACCGGCTGCTCCGAGTCTTTCGAG	114
DB	53	CTGCCCTGCGCACAGGACATCAACCCCTACCGGTGTGGATCGGAATCATGCTGCAG	112
QY	115	CAGACCCGGGTGAGAGAGCCCTCCCTTACCCGCGCTTCTGAGGCGTTCCAC	174
DB	113	CAGACCCAGGTAGACACCTGCTCGTTACTTGACCGTTTATGACAGCTTGCCGAC	172
QY	175	CTGAAGCCCTGCGCGCGCTTCTCCGAAAGGTCCTTAGGCTTCGAGGGGCGGG	234
DB	173	GTCGAGGCACTGCGCGCGCGCGCGCGAGCAGAGTCTGCACCTGTCGACCGGGCTCGGC	232
QY	235	TACTACCGCGCGGAGACCTTCCACCGCTGCGCCGACAGGT-----G	279
DB	233	TACTACAGCGCGTGGCGGCAACTGACAGAACCGCGCAATGTGTGACAGCGCATGCG	292
QY	280	GAGGAGCTTCCCGCGAGCTTCCGCGGAGCTTCTGAGTCTCGGGCTTACAC	339
DB	293	GAGGAGTTCCTCCCGGAGCTTCCGCACTGCGCACTCGCCGATCGGCGCTCACC	352
QY	340	GCGGCGCGTGGCTTCATCGCTTTCGAGGAGCGGTCGAGCGGTGACGGGAAGTC	399
DB	353	GCTGAGGCGATCGGCGCTGTCATGAGGCTGCGGCGACCGATCTTCAGCGCAAGTC	412

QY 400 CCGAGGCTCTCTCCCGCCTCTTGCCTCCCGGAAAGCCCAAGAGAGAGCTTTTGGCC 459  
DB 413 AAGCGGTACTGTCGCGCTACTTGGCGAGAGACGGCTATCCCGGCAACGAAAGTGGCC 472  
QY 460 CTGGCCAGAGGCTCTCTCTCCCGAGGCGTGGACCCGCGGGGTTGG-----AAC 507  
DB 473 AGGGCGTGTGGAAAGCCCGCAAGCTTCACTCCCGACGACGGGTCAACCTACAC 532  
QY 508 CAGGCTCATATGAGCTCGGGGCGACGCTCTGCTCCGAAAGCGCCCGCTTGGCGGGCC 567  
DB 533 CAGGCAATGATGACCTTGGGGCCACCTCTGACACGCGAGCAAGCCGAGTTGCTCTT 592  
QY 568 TGGCCCCCTAGAGGCGCTTCTGCGCGGGAA 596  
DB 593 TGGCCGCTGTCTCGGCTGCGCGCGCA 621

## RESULT 2

US-09-252-991A-6886/c  
Sequence 6886, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR APPLICATION NUMBER: 1999-02-18  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6886  
LENGTH: 1401  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6886

Query Match Best Local Similarity 12.9%; Score 125.4; DB 4; Length 1401;  
Matches 311; Conservative 0; Mismatches 231; Indels 27; Gaps 2;

QY 55 CTCCCTGCGGGGAGAGAGACCTTACCGCGTCTGCTTCCGAGTCTTTCGAG 114  
DB 1137 CTGCTTGGCAGACAGAGGATACCCCTTACCGGGTGTGGTCTCGAAATCATGCTGAG 1078  
QY 115 CAGACCCGAGTGAAGAGAGCCCTTCCCTATTACCGCGCTTTCGAGAGCGCTTCCACC 174  
DB 1077 CAGACCTCAAGTACAGACCGTGTCTGCTTACCTTGAACGCTTTCAGAGCGTTCGAC 1018  
QY 175 CTGAAGGCTGCGCGCGCTTCCCTGGAAGAGTCTTAAAGGTCTGAGAGGAGCGGCG 234  
DB 1017 GTGAGGAGACTGCGCGCGCGCGCGAGAGAGTCTTCACTGTGAGACCGGGCTCGGC 958  
QY 235 TACTACCGGCGGGGAAACCTTCAACCGCTGCGCGCAAGCGT-----G 279  
DB 957 TACTACAGCGGTGTGCGCAACCTGCAAGACCGCGCAAGTGTGTGAGAGCGAGTGC 898  
QY 280 GAGGAGCTTCCCGAGCTTGCAGAGCTTTCGAGAGCTTCTGTGCTGAGAGCTTACAC 339  
DB 897 GGGGAGTTCCTCCCGACGTCGAGCACTGCGCAAGTCCCGGATGCGCGCTCAC 838  
QY 340 GCGGCGGAGTGGCTTCCCTTCCGAGAGCGGAGTGGCGGAGTGGAGCGGAAAGTC 399  
DB 837 GCTGAGAGCATGCGCAAGCTGTGATGAGGCTGCGCGACGACCATCTCGAGGAGAGTC 778  
QY 400 CCGAGGCTCTTCTCCGCTCTTTCGCGCGGAAAGCCCAAGAGAGAGCTTTTGGCC 459  
DB 777 AAGCGCTATGAGCTGCGCTTACCTGCGAGAGAGCGCTATCCCGCAACCGAAGTGGCC 718  
QY 460 CTGCGCCAGGCGCTCTCTCCGAGAGCGGTGAGCCCGGAGTGTGG-----AAC 507

DB 717 AGGGCGTGTGGAAAGCCCGCAAGCTTCAACCCCGACGACGGGTCAACCTACAC 658  
QY 508 CAGGCTCATATGAGCTCGGGGCGACGCTTGGCTTCCGCAAGAGCCCGCTTGGCGGGCC 567  
DB 657 CAGGCAATGATGACCTTGGGGCCACCTCTGACAGCGAGCAAGCCGAGTTGCTCTT 598  
QY 568 TGGCCCCCTAGAGGCGCTTTCGCGGGGAA 596  
DB 597 TGGCCGCTGTCTCGGCTGCGCGCGCA 569

## RESULT 3

US-09-252-991A-6739  
Sequence 6739, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR APPLICATION NUMBER: 1999-02-18  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6739  
LENGTH: 1413  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6739

Query Match Best Local Similarity 12.9%; Score 125.4; DB 4; Length 1413;  
Matches 311; Conservative 0; Mismatches 231; Indels 27; Gaps 2;

QY 55 CTCCCTGCGGGGAGAGAGACCTTACCGCGTCTGCTTCCGAGTCTTTCGAG 114  
DB 409 CTGCTTGGCAGACAGAGGATACCCCTTACCGGGTGTGGTCTCGAAATCATGCTGAG 468  
QY 115 CAGACCCGAGTGAAGAGAGCCCTTCCCTATTACCGCGCTTTCGAGAGCGCTTCCACC 174  
DB 469 CAGACCTCAAGTACAGACCGTGTCTGCTTACCTTGAACGCTTTCAGAGCGTTCGAC 538  
QY 175 CTGAAGGCTGCGCGCGCTTCCCTGGAAGAGTCTTAAAGGTCTGAGAGGAGCGGCG 234  
DB 529 GTGAGGAGACTGCGCGCGCGCGAGAGAGTCTTGAACCTGTGAGACCGGGCTCGGC 588  
QY 235 TACTACCGGCGGGGAAACCTTCAACCGCTGCGCGCAAGCGT-----G 279  
DB 589 TACTACAGCGGTGTGCGCAACCTGCAAGACCGCGCAAGTGTGTGAGAGCGAGTGC 648  
QY 280 GAGGAGCTTCCCGAGCTTGCAGAGCTTGGAGGCTTCTGTGCTGAGAGCTTACAC 339  
DB 649 GGGGAGTTCCTCCCGACGTCGAGCACTGCGCAAGTCCCGGATGCGCGCTCAC 708  
QY 340 GCGGCGGAGTGGCTTCCCTTCCGAGAGCGGAGTGGCGGAGTGGAGCGGAAAGTC 399  
DB 709 GCTGAGAGCATGCGCAAGCTGTGATGAGGCTGCGCGACGACCATCTCGAGGAGAGTC 768  
QY 400 CCGAGGCTCTTCTCCGCTCTTTCGCGCGGAAAGCCCAAGAGAGAGCTTTTGGCC 459  
DB 769 AAGCGCTATGAGCTGCGCTTACCTGCGAGAGAGCGCTATCCCGCAACCGAAGTGGCC 828  
QY 460 CTGCGCCAGGCGCTCTCCCGAGGCGGTGAGCCCGGAGTGTGG-----AAC 507  
DB 829 AAGCGCTATGAGAGCGCGCAAGCTTCAACCCCGACGACGAGTCAACCTACAC 888  
QY 508 CAGGCTCATATGAGCTGCGGCGACGAGTGTGCTGCGAAAGCGCCCGCTTGGCGGGCC 567  
DB 889 CAGGCAATGATGACCTTGGGGCCACCTCTGACAGCGGCAAGCCGAGTTGCTCTT 948

Qy 568 TGCCCCCTAGGGGCTTCTGCGGGGAA 596  
Db 949 TGCCCGCTGTCTCCGGCTGCGCGGCA 977

## RESULT 4

US-09-489-039A-4788  
; Sequence 4788, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709, 2004001  
; CURRENT APPLICATION NUMBER: US/09/489, 039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 4788  
; LENGTH: 1551  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-4788

Query Match 11.6%; Score 113; DB 4; Length 1551;  
Best Local Similarity 53.0%; Pred. No. 9.2e-15;  
Matches 307; Conservative 0; Mismatches 245; Indels 27; Gaps 2;

Qy 53 CCTCCCTGCGGGGGAAGAACCTTACCGCTGCTGCTCCGAGTCTCTTCG 112  
Db 560 CCTGCGCTGGCAAAATCGCCAGAGCCCTTACAAATATAGGCTCTCGAAGATGTTTCG 619  
Qy 113 AGCAGACCGGGGTGAGCAGCGCCCTCTTATACCGCGCTTTCTGAGCGCTTCCCA 172  
Db 620 AACCAACCCAGGTGACACAGGTATCCCTATTGTAACGCTTTATGAGCGCTTCCCA 679  
Qy 173 CCTGAGGCGCTGCGCGCGGCTTCCCTGAGAGGCTCTTAGGCTCTGAGCGGGGGG 232  
Db 680 CGGTGTGATCTCGCCACAGCGCGCGCTGATGATGCTGATCTGTGACCGGCTTCG 739  
Qy 223 GCTACTACCGCGGGGGAACAC-----CTCCACCGCTGCGCGCGGAGCG 277  
Db 740 GCTACTACCGCGGGGGAACACTTACATTAAGCGCGGAGCAAGTGGCCACACTGCAAG 799  
Qy 278 TGGAGAGCTTCCCGAGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAGCTTCA 337  
Db 800 GCGGGGAATTCCTCCGAGCTTTCAGAGAGTGGCGCGCTACCGCGCGCTCGGCGCTCA 859  
Qy 338 CGCGCGCGCGGCTGCGCTCATCGCTTGGGGAGCGGGTGGCGGGGTGAGCGGAAG 397  
Db 860 CGCGCGCGCGGATTTTATCTCTTGGCTGCTGATTCGATTCGATTCGAGCGCAAG 919  
Qy 398 TCCGAGAGGCTCTCTCCGCTCTTCCGCGCGGGAAGCCCAAGAGAGAGCTTTTCG 457  
Db 920 TGAAGCGCTGCTCCCGCGCTGCTATGCTGACGCGCTGCGCGGGGAAAAAGAGTGG 979  
Qy 458 CCTGCGCCAGGCGCTCTCTCCCGAGGCGTGAAGCCCGGGGCTGTG-----A 505  
Db 980 AAAAAGCGCTGGGACATCAGCAAGAGGTCAACCCCGGAGAGGGCGTGAAGCGTTCA 1039  
Qy 506 ACCAGGCGCTCATGAGCTCGGGGCAAGGTCTGCTGCGGAAAGCGCCCGTTGGGGG 565  
Db 1040 ACCAGGCGATGATGATCTCGGGGCAATGATTTGCAACCGCTCGAAGCGGAAGTGGAGC 1099  
Qy 566 CTGCGCCCTAGAGGCTTCTGCGCGGGGAGAGAGGCC 604  
Db 1100 TGTGTCTGCTGAGCAAGCGCTGCTGCTTACGCAACC 1138

RESULT 5  
US-09-252-991A-6783  
; Sequence 6783, Application US/09252991A

; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252, 991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074, 788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094, 190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6783  
; LENGTH: 549  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6783

Query Match 11.0%; Score 107.2; DB 4; Length 549;  
Best Local Similarity 54.4%; Pred. No. 1.3e-13;  
Matches 248; Conservative 0; Mismatches 193; Indels 15; Gaps 1;

Qy 55 CTCCCTGCGGGGGAAGAACCTTACCGGCTGCTCTCCGAGTCTCTTCGAG 114  
Db 60 CTGCGCTGCGAGAGGAGCATCAACCCCTACCGGGTGTGGGTCTCGAATCATGCTGCA 119  
Qy 115 CAGACCGGGGTGAGCAGCGCTCTCCCTTATTAACGCGCGCTTCTGAGCGCTTCCACC 174  
Db 120 CAGACCCAGGTGAGCAGCGCTCTCGGTACTTCACTTCACTGACAGCGTTGCCGAC 179  
Qy 175 CTGAGGCGCTGCGCGCGGCTTCCCTGAGAGGCTCTTAGGCTCTGAGCGGGGGG 234  
Db 180 GTGAGGCACTGCGCGCGCGCGCGCGAGAGGCTCTGACCTGTGAGCGGGCTCGGC 239  
Qy 235 TACTACCGCGGGGGAACACCTTCAACCGCTGCGCGGAGCT-----G 279  
Db 240 TACTACAGCGCTGCGCGGCACTGACAGAGCGGCAATGCTGTGACAGCGGATGCG 299  
Qy 280 GAGAGCTTCCCGAGCTTCCCGAGCTTCCGAGGCTTCTGCTGCTGCTTCAACC 339  
Db 300 GGGGAGTTCCCGCGGAGCGTGAAGCACTGCGGAATCCCGGCACTCGGCGCTCCACC 359  
Qy 340 GCGCGCGGCTGCTTCCATTCGCTTCCGAGAGCGGGTGGCGGCTGAGCGGAAGCTC 399  
Db 360 GCTGAGCGCATGCGCAGCTGTGATGAGCTGCGGAGCGCATCTTCAAGCGCAAGCTC 419  
Qy 400 CGAAGGCTCTCTCCGCTTTCGCGCGGGAAGCCCAAGAGAGAGCTTTTCGCC 459  
Db 420 AAGCGCTACTGCGCGCTTCACTGCGCGAGAGCGGCTATCCCGGCAACGAAAGTGGC 479  
Qy 460 CTGCGCCAGGCGCTCTCTCCCGAGGCGTGAACCG 495  
Db 480 AGGGCGCTGTGGAGAGCGCGCAAGCTTCAACCGG 515

RESULT 6  
US-08-813-574-1  
; Sequence 1, Application US/08813574  
; Patent No. 6013473  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Ying-Fei  
; TITLE OF INVENTION: Human Muty  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/813,574
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/013,132
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1858 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-813-574-1

```

```

Query Match      10.7%; Score 104; DB 3; Length 1858;
Best Local Similarity 54.0%; Pred. No. 6,7e-13;
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

```

```

69 GGAAGAGACCTTACCGGCTCTGATCCGAGGTCCTTTCGACGACACCCGGGTGA 128
498 GAGACAGCGGGCAATAGCTGTGGGTCTCAGAGGTCAATGCTGACGACACCCAGTTGC 557
129 GCAGGCGCTTCCCTATTACCCGCCCTTTTGTGAGCGCTTTCCACCGAAGGCGCTTGC 188
558 CACTGTGATCACTACTATACCGATGATGCAAGATGGCCCTTCACTGACGAGACTTGGC 617
189 CGGGGCTTCCCTGGAAGAGGTCCTTAGGCTGCGCAAGGGGCGGGCTACTACCGCGGGC 248
618 CAGTGTCTCCCTGAGAGAGGTGAATCACTCTGGGCTGGCGCTACTATTCTCTGTGG 677
249 GGAACACTCCACCGCTGCGCCGGAAGCTGAGAGGCTTCCCGGAGCTTC----- 300
678 CCGGCGGCTGACGAGAGGAGCTCGAAGGTGTAGAGAGCTTAGGGGCGACATGCGACG 737
301 -----GCGAGCTTGGGGGCTTCTGTCTCGGCTTACACCGCGCGCGGT 350
738 TACAGCAGAGACCTTGCAGACGCTCTGCTGGGCTGGGCGCTACACAGCTGGGCGCAT 797
351 GGCTTCATCGCTTTCGGGAGCGGGTGGCGGCTGGAAGGAAAGTCCGAGAGGCTCT 410
798 TGCTCTATCGCTTTGGCCAGGCAACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 857
411 CTCCTGCTTTCGCCC-----GGGAAAGCCCAAGAGAGAGCTTTTTCGCTTGCCTG 468
858 GTGCGGTGTTCGAGCATTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 917
469 GGCCTCTTCCCGAGAGGCTGAGCCG-----GGGTGTGGAACAGAGGCGCTCAT 518
918 GGGTCTAGCCGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 977
519 GGAAGCTCGGGGCAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 578
978 GGAAGCTAGGGGCAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1037
579 GGCCTTGTGCGGG 592
1038 GAGCTGTGTGCGGG 1051

```

RESULT 7

```

US-09-651-656-38
Sequence 38, Application US/09651656
Patent No. 6340566
GENERAL INFORMATION:
APPLICANT: MCUTCHEN-MALONEY, SANDRA
APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,
TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES
FILE REFERENCE: IL-10689
CURRENT APPLICATION NUMBER: US/09/651,656
PRIORITY FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 60/192,764
PRIOR FILING DATE: 2000-03-28
NUMBER OF SEQ ID NOS: 106
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 2293
TYPE: DNA
ORGANISM: Escherichia coli
US-09-651-656-38

```

```

Query Match      8.9%; Score 87.2; DB 3; Length 2293;
Best Local Similarity 54.1%; Pred. No. 1.9e-09;
Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;

```

```

53 CCTCCCTGCGGGGAGAGACCTTACCGCGTCTGCTCTCCGAGGTCCTTCTGC 112
1161 CTCTGCTTGGCAATTGACAGAGCGCTTACAAAGTATGCTTACAGATGATTTGC 1220
113 AGCAGACCGGGGTGAGAGCGCCCTCCCTATTACCGCGCTTTCGAGCGCTTCCCA 172
1221 AACAACTCAGGTGCGACCTTATCCCTATTGTAAGCTTTATAGCGCGCTCCCA 1280
173 CCTGAGACCGCTGCGCGCTTCCCTGGAAGGTCCTTAGGTCGTGCAAGGGGCGG 232
1281 CGGTATGCGATTCGCGCAATGCGCGCTCGAGAACTTCTCACTGTGACCGGCTTG 1340
233 GCTACTACGGGCGGCGGACACCTCCACCGCTGCGCGGAGCGTGAAG----- 282
1341 GCTATTACCGCGCGCGCAATCTGTCAATTAAGCGGACAAAGTGCGGCACTTACACG 1400
283 -----GAGCTTCCCGGAGCTTTCGCGAGCTTTCGGGGCTTCTGTGTGCGGCTTACA 337
1401 GCGGTAAATTCGCGGAACTTTTGAAGAGTTCAGACCTCCGCGCTCGGGCTTCCA 1460
338 CCGCGGCGGCTGAGCTTCATCGCTTTCGGGAGCGGAGTGTGCGGCGGTGACGGAACG 397
1461 CCGAGGCGGAGTCTCTGCTTCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1520
398 TCGGAGAGTCTCTCCGCTTTCGCG 425
1521 TCAAAGCGTGTGCGGCGGCTGTATGC 1548

```

```

RESULT 8
US-09-650-855-38
Sequence 38, Application US/09650855
Patent No. 6365355
GENERAL INFORMATION:
APPLICANT: MCUTCHEN-MALONEY, SANDRA
APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA
MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA
TITLE OF INVENTION: MISMATCHES
FILE REFERENCE: IL-10284
CURRENT APPLICATION NUMBER: US/09/650,855
PRIORITY FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 60/192,764
PRIOR FILING DATE: 2000-03-28
NUMBER OF SEQ ID NOS: 106
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38

```

LENGTH: 2293  
TYPE: DNA  
ORGANISM: Escherichia coli  
US-09-650-855-38

Query Match 8.9%; Score 87.2; DB 3; Length 2293;  
Best Local Similarity 54.1%; Pred. No. 1.9e-09;  
Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;

QY 53 CCTCTCCCTGGCGGGGAGAGAACCTTACCGCGTCTGCTCCGAGCTCTTCTGC 112  
DB 1161 CTCTCCCTGGCGAAATTGACAAAGACGCTTCAAGATAGCTTCAGAGATGTTTC 1220  
QY 113 AGCAGACCGGGGTGAGAGACGCTTACCGCGTCTGCTCCGAGCTCTTCTGC 172  
DB 1221 AACAACTCAGGTGTGACACCGTTATCCCTATTTTGAACGCTTTATGCGCGCTTCCCA 1280  
QY 173 CCTGAGAGCCCTGCGCGGCTTCCCTGGAAGAGTCTTGAAGTCTGCGAGGGGCGG 232  
DB 1281 CGGTGACCGATCTCGCATAGGCGCGCTGAGAGATTCACATTGTGACCGGCTTG 1340  
QY 233 GCTACTACCGGCGGGGGAACCTTCACCGCTGCGCGGCGGAGCGTGAG----- 282  
DB 1341 GCTATTACGCGCGCGCGCAATGCAATAAGCGGCAACAAGTGCGGACCTTACAG 1400  
QY 283 -----GAGCTTCCCGGAGCTTCGCGAGCTTCGGGGGCTTCTGTCTCGGCGCTTACA 337  
DB 1401 GCGGTAATTCGCGAATCTTGAAGAGTTGACGACATCGCGGCGCTTCA 1460  
QY 338 CGCGCGCGCGGTGCGCTTCATCGCTTCGGGAGCGGTGCGGCGGTGAGCGGAGCG 397  
DB 1461 CGCGAGGCGGATCTCTCGCTTCTCTGGTAAAGACTTCCGATTCTCGACGGTACG 1520  
QY 398 TCCGAGGGTCTCTCTCGCGCTTCTTTCG 425  
DB 1521 TCAACGCGTCTGCGCGCTGTATGC 1548

## RESULT 9

US-09-103-840A-2  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-2007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 8.9%; Score 86.6; DB 3; Length 4403765;  
Best Local Similarity 50.2%; Pred. No. 6.3e-09;  
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

QY 22 CTCCTCGCTGTGTACCGGAAACCGCGCGCTTCCCTGCGGGG---GAGAAAGAC 78  
DB 4022631 CTTCTCGCTGTGTATCAGGATCGACCGGATCTGCCCTGCGAGAGCCCGGTGTCAAC 4022690  
QY 79 CCTTACCGGCTGTGTCTCGAGGATCTTCTGAGAGACCGCGGTTGAGAGAGCCCTC 138

DB 4022691 CCGTGGCAGATCTGTGTACGAGATTACGTGAGACACGCGCGCGCGGGTGTG 4022750  
QY 139 CCTTATTACCGCGCTTCTTGTGAGGCTTTCCACCTTGAAGCCCTGCGCGGCTTCC 198  
DB 4022751 GCGATCTGCGCGAGCTGTGGGTGGGCGTGGCCACGCGCTTCCGCCACCGGACAG 4022810  
QY 199 CTGGAAGAGTCTTAAAGGTCTGAGAGGGGGGGGTACTACCGGAGGGGCGGAACCTTC 258  
DB 4022811 ACCGCGGATGTATACGCGCTTGGGCAAGCTGGGCTTATCCAGAGAGCAAGCTTCA 4022870  
QY 259 CACCGCTGCGCGCGAAGCGTG-----GAGAGCTTCCCGAGCTTTCGCG 303  
DB 4022871 CACGAGTGGCCACCGGTATACCGCGCGACCAACATGAGTGTGCGACAGTATCGAG 4022930  
QY 304 GAGCTTGGGGGCTTCTGTGTCTGCGGCTTACACCGCGCGCGGTGCTTCATCGCC 363  
DB 4022931 ATCTGTGTACCTCTGCGCGCGCTGCGGAGCTACCGCGCGCGGTGCGTTCGCT 4022990  
QY 364 TTGCGGAGAGGGGTGCGCGGCTGAGACGGGAACGTCCGAGGGTCTTCCGCGCTT- 422  
DB 4022991 TACCGCGAGCGGTGCGGTGTGACACCAATGTGCGGTGTGCGCGCGCTT 4023050  
QY 423 -----GCGCGGGAAGCCCAAGAGAAAGACTTTCGCTTCCGCGCGAGCG 471  
DB 4023051 CACGCGCGCGCGACCGCGGTGCGGTGCGGTGCGCGCGCGACGACGCTTGTGCG 4023110  
QY 472 CTCCTCCCGAGAGGTGAGACCGGGGCTGTGAACCAAGCCCTCATGAGCTCGGGGCC 531  
DB 4023111 CTGTTCGCGACCGGAGAGCGCGCTGTAATTTGCGTGCCTGTATGAGTGTGGTG 4023170  
QY 532 AGGTCTGTGCTGCGCAACCGCGCTTGGGGGCTGCGCGCTTGAAGGCTTCTTCGCG 591  
DB 4023171 AGGTGTGACCGCGCGCACCGCGGTGCGGTATGCGCGTGAATGCTGTGCGATG 4023230  
QY 592 GGAAGAGAGCGCGCGCGCTACCGCGCGCGAGAAAGCGCGGCGGAAGAGAGCGCG 651  
DB 4023231 CGGATGCGCGGTATTCGCGCGCTGAGACGCTC---GCGCGCGCGGCGGCGCTACAC 4023287  
QY 652 CTCGTGCGCTCTCTCTCTCTCGGGCGG 678  
DB 4023288 GGAACCGACCGCAAGTCCGCGGACGG 4023314

## RESULT 10

US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-2007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 8.9%; Score 86.6; DB 3; Length 4411529;  
Best Local Similarity 50.2%; Pred. No. 6.3e-09;  
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

QY 22 CTCCTCGCTGTGTACCGGAAACCGCGCGCTTCCCTGCGGGG---GAGAAAGAC 78

```

Db 403050 CTTCTCGCTTGGTATACAGGATGACACCGGATCTGCTTGGCGAGACCCGGGTGTACG 4030609
Qy 79 CTTTACCGGCTCTGTGCTTCGAGGTCTTTCGACGACACCGGGGTGAGACCGCTTC 138
Db 4030610 CCGTGACAGATCTGTGTACAGCGATTCATGTGACAGACACCGCCGGCCGGTGTGTG 4030669
Qy 139 CCGTATACCGCGCTTTCGAGCGCTTTCGACCGCTTGAAGCGCTTGGCCGGCTTC 198
Db 4030670 GCGATTCGCGAGCTGGGTGGCGCGGTGGCCACCGCTTCGACACCGCCGACG 4030729
Qy 199 CTGAGAGAGTCTTTCGAGGTCTTCGACGAGGCGCGGCTACACCGCGCGGAGACCTTC 258
Db 4030730 ACCGCCATGTGTACCGCGCTGGGCGACATGAGCTATCCAGAGGACCAAGGCTTA 4030789
Qy 259 CACCGCGCTGGCGGAGGCTG-----GAGAGCTTCCCGGAGCTTTCGCG 303
Db 4030790 CACGATGCGCACCGCTATGCGCCCGGACCAACATGACGTGGTGCCTCGATATGAG 4030849
Qy 304 GAGCTTCGGGGGCTTCTGTGCTTCGAGCTTACACCGCGCGCGGCGGCTTCATCGCC 363
Db 4030850 ATCTGTGACCGCTGCGGGCGGTGCGGAGCTACACCGCGCGCGGTGGCTTTCGCT 4030909
Qy 364 TTCGGGAGCGGCTGCGCGGCTGAGCGGAGACGTCCGAGAGGTCTTTCGCGCTTC 422
Db 4030910 TACCGCGAGCGGCTGCGCGGTGAGCACATGTGCGCGCGGTGCGCGCGCTT 4030969
Qy 423 -----CGCGCGGAGAACCCCAAGAGAGAGCTTTTGCCTTCGCGCGAGGC 471
Db 4030970 CACGCGCGCGCGACCGCGGTGCGCTATGCTGCGCGCGACACGATTTTGGCG 4031029
Qy 472 CTCCTCCCGAGGGGCTGAGACCGGGGCTGTGAAACAGGCTTCATGAGCTCGGGCC 531
Db 4031030 CTGTTCGCCGACCGGACGACGCGCTGAAATTTTCGTGCGCGCTGATGAGATGGGTGCG 4031089
Qy 532 ACCGTTCGCTTCGGAAGCGCGCTTTCGCGGCGCTGCGGAGCTTCGCGCG 591
Db 4031090 ACCGTGTGACCGCGCGACACCGCGGTATGCGGCTTATGCGCTGAGCTGTGGCATG 4031149
Qy 592 GGGAGAGAGCGCGCGCGGCTACCGCGCGCGCGAGAGCGCGCGGAGAGAGCGC 651
Db 4031150 CGGATTCGCGGTATTCGCGCGCTGAGAGGTCC--GCGCGCGCGGGGAGGCTTACAC 4031206
Qy 652 CTCGTTCGCGCTTCGTCTCTTCGCGCGCG 678
Db 4031207 GGAACCGACGCGCAAGTCCGCGGACGG 4031233

```

RESULT 11  
US-09-583-110-1922  
; Sequence 1922, Application US/09583110  
; Patent No. 6699703

GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 1922

LENGTH: 1176  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae

US-09-583-110-1922

Query Match 8.4%; Score 82; DB 4; Length 1176;  
Best Local Similarity 52.2%; Pred. No. 2,1e-08;

```

Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1
Qy 22 CTCCTCGCTGATACCGGAAAAACGCGCCCTCTCCCTGGGGGAGAGAGACCT 81
Db 70 CTCCTCGCTGATGATGAAAAAAGAAATTTGCTTGAAGAGAACTAAAAATCT 129
Qy 82 TACCGGTCTGTCTCGAGAGCTCTTCGACGACACCGGGGTGAGAGAGCCCTCC 141
Db 130 TATCAATCTGGTATCTGAAATCATGCTTCAGACGACAGAGGTGATACATGTTATCT 189
Qy 142 TATTACCGCGCTTTCGAGAGCTTTCACACCTGAAAGCGCGCGGCTTCCCTG 201
Db 190 TATTAGAAAGATCTTGAGCTGTTTCCACTGTGAAATCTGGCAATCGCCCTGAG 249
Qy 202 GAAGAGTCTTATAGGTCTGCGAGGGGCGGCTA-----CTACCGCGG 246
Db 250 GAGAGTTTACTGAAAGCTTGGAGGCTTGGCTTATTTCTCGAGTTCCGAAATATCAG 309
Qy 247 GCGGACACCTCCACCGCTGCGCGGAGAGCTTCCCGAGGCTTCGCGAG 306
Db 310 GCTGACCGCAGAGATATGACTGACTTGTGTGCAATTTCCAAATACCTATGAGGA 369
Qy 307 CTCGCGGCTTCTGTCTTCGCGGCTTACACGCGCGCGGTGCGCTTCATCGCTTC 366
Db 370 ATTTCAGCTTGAAGGATTTGACCTTACACGACGAGAGCCATTTCCAGTATCTTT 429
Qy 367 GGGGAGCGGTGCGCGGTGAGCGGAGACGTCCGAGGCTCTCCCGGCTTTCG 424
Db 430 AACTTCCTGAGCAGCTGTAGATGTATGTATGTCAGGTGGGTCTTGGCGCTCTTTG 487

```

RESULT 12  
US-08-961-527-12/c  
; Sequence 12, Application US/08961527  
; Patent No. 6420135

GENERAL INFORMATION:  
; APPLICANT: Charles Kunesh  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9909 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

US-08-961-527-12

Query Match 8.4%; Score 82; DB 4; Length 9909;  
 Best Local Similarity 52.2%; Pred. No. 2,7e-08;  
 Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

QY 22 CTCCTGCTGTACCGGAAAAAGCGCCCTCCCTGCGGGGAGAGACCTT 81  
 DB 5808 CTCCTGCTGTGTATGATGAAAAAAGATTTCTTGGAGAGAGATTAATAATCTT 5749  
 QY 82 TACCGGCTGCTGTCCGAGGCTTCTGAGAGACACCGGCTGAGACAGCGCTCC 141  
 DB 5748 TATCAATGTGGATATCTAAATCATGTTCAGACACAGAGGTGATACAGTTATCC 5689  
 QY 142 TATTACCGCGCTTCTGAGAGCGCTTCCACCTGAGAGCCCTGAGCGGCTTCCCTG 201  
 DB 5688 TACTAGAAAAAGATTTTGGACTGTGTTCCACTGTGGAAGTGTGCACTGCGCTGAG 5629  
 QY 202 GAAGAGTCTTAAAGGTCTGGCAAGGGGCGGGCTA-----CTACCGGCGG 246  
 DB 5628 GAGAGTTTACTAAAGCTTGGAGGGGCTTGGGCTATTATTCTGAGTTGCAATATGCA 5569  
 QY 247 GGGGAACACTTCCACCGCGCTGCGCCGAGCGTGAAGAGCTTCCCGAGCTTCCCGAG 306  
 DB 5568 GCTGAGCCCAAGACATTAATGACTGACTTGTGGCCAAATTCATAATGTAAGA 5509  
 QY 307 CTTCGGGGGCTTCTGTCTCGGGCTTACACCGCGCGCGGTGCGCTCAATCGCCTTC 366  
 DB 5508 ATTTCAGCTTAAAGAGGATTTGACCTTACAGCAGAGACATTTCCAGTATGCTTTT 5449  
 QY 367 GGGAGCGGGTGGCGCGGTGACGCGGAACGTCCGAGGAGCTCTCCGCGCTTTCG 424  
 DB 5448 AACTTGCTGAGCAGCAGCTAGATGTGTAATGTCATCGGGGTCTGCGCGCTGTTTG 5391

RESULT 13  
 US-09-864-866-46  
 ; Sequence 46, Application US/09864866  
 ; Patent No. 6723548  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lloyd, R. Stephen  
 ; APPLICANT: McCullough, Amanda K.  
 ; APPLICANT: Nguyen, Khia  
 ; TITLE OF INVENTION: DNA REPAIR POLYPEPTIDES AND METHODS OF USE  
 ; FILE REFERENCE: 265, 00170101  
 ; CURRENT APPLICATION NUMBER: US/09/864,866  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/206,279  
 ; PRIOR FILING DATE: 2000-05-23  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 46  
 ; LENGTH: 968  
 ; TYPE: DNA  
 ; ORGANISM: Micrococcus luteus  
 US-09-864-866-46

Query Match 8.3%; Score 81.2; DB 4; Length 968;  
 Best Local Similarity 47.2%; Pred. No. 3e-08;  
 Matches 300; Conservative 0; Mismatches 318; Indels 18; Gaps 1;

QY 44 AGCCCGCCCTCCCTGCGGGGAGAGAACCTTACCGCGCTGCTCCGAG 103  
 DB 200 ACCCGTACCGCTTCCCGAGCTTGAAGACCGCTTCAAGTCTGTCGACCG 259  
 QY 104 TCTCTTCTGACAGACCCCGGTGAGACAGCCCTCCCTATTATACCGCCCTTTGAGAC 163  
 DB 260 TCTGTCCGCGCCAGACCAAGCAGTGCCTGCAAGCAGCACAGCGCGGCTGTTGCC 319  
 QY 164 GCTTTCACACCTGAAAGGCTTGGCGCGCTTCCCTGGAAGAGTCTTAAAGGTGTC 223  
 DB 320 GCTTCCCGAGTCCACGAGATGGCGCGCCACGAGCCGAGCTGCGAGAGCTGTGC 379  
 QY 224 AGGGGCGGGCTACTACCGG-----CGGGCGAACAACCTTCAACGCGC 265

DB 380 GCTCCAGGGGTTCTACCGGAACAAGCGCTCCGAGATCTGCGGCTGTCCAGAGCTCG 439  
 QY 266 TGGCCCGAAGCGTGAAGAGCTTCCCCAGAGCTTCCCGAGCTTCCGGGCTTCTGTGTC 325  
 DB 440 TGGCCCGGACAGACCGCGAGGTCCCGCCGCTTCTGAGAGACTGTGTGGGCTGCGGG 439  
 QY 326 TCGGGCTTACACCGCGGGGCGGTGGCTTCCATCGCTTGGGAGAGCGGTGCGGCGG 385  
 DB 500 TGGCCCGAAGACCGGCTTGTGTGTCTGCGCAAGCGCTTCCGCGAGCCCGGATCAC 559  
 QY 386 TGAACGGAACGTCCGAGAGGTCTCTCCGCTTCTTCCCGGGAAGCCCCAAGAGA 445  
 DB 560 TGAACAGCACTTCCGCGCGCTGCGCGCGCTCGGGGTTACAGAGACGAGACCGAC 619  
 QY 446 AGAGCTTTGCGCTTCCGCGCGAGGCTTCTCCCGAGAGGCGTGAACCGGGGCTGGA 505  
 DB 620 GTAAAGGTGACACCGCTGGGCGCTGTTCCCGCGGAGCTGACATCTCT 679  
 QY 506 ACCAAGCCCTCATGAGCTCGGGGCGCAGGCTGCTGCGGAACCGCCGCTTGGCGGG 565  
 DB 680 CCGACCGGCTGATCTTCCACGCGCGCGCTGTGCTCAGCGCGCGCGCGCTGCGGG 739  
 QY 566 CCGCCCGCTAGGGGCTTCTGCGGGGAGAGAGCGCCCGGCGCTACCCCGCGCA 625  
 DB 740 GGTGCCGATCGCGCGCTGTGCGCTCTACGCGCGGGGAGACGACCCGAGCGGG 799  
 QY 626 GGAAGCGCGCGGGAAGAGAGCGCTGTGCGCC 661  
 DB 800 CGCGCGCTGCTGCGCTACGAGCTCAACCGCGGC 835

RESULT 14  
 US-09-107-532A-3654  
 ; Sequence 3654, Application US/09107532A  
 ; Patent No. 6583275  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 7310  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 ; STREET: 100 Beaver Street  
 ; CITY: Waltham  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02354  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-ROM ISO9660  
 ; COMPUTER: PC  
 ; OPERATING SYSTEM: <Unknown>  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/107,532A  
 ; FILING DATE: 30-Jun-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/085,598  
 ; FILING DATE: 14 May 1998  
 ; APPLICATION NUMBER: 60/051571  
 ; FILING DATE: July 2, 1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ariniello, Pamela Deneke  
 ; REGISTRATION NUMBER: 40,489  
 ; REFERENCE/DOCKET NUMBER: GTC-012  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (781)893-5007  
 ; TELEFAX: (781)893-8277  
 ; INFORMATION FOR SEQ ID NO: 3654:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1212 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...1212  
SEQUENCE DESCRIPTION: SEQ ID NO: 3654;  
US-09-107-532A-3654

Query Match 6.5%; Score 63.6; DB 4; Length 1212;  
Beef Local Similarity 47.6%; Pred. No. 0.0013;  
Matches 278; Conservative 0; Mismatches 279; Indels 27; Gaps 2;

31 TGTATCGGAGAAAGCGCCGCTCCCTGCGGGGAGAAAGACCTTACCGCGT 90  
100 TGGTATGAAACAAAGAAAGAAACCTGCTTGGCGCTATATGCTATCTTATCGTATC 159  
91 CTGATCTCGAGGCTCTTCTGACAGACCGGGTGAAGAGCCCTCCCTATTACCGC 150  
160 TGGATCTCGAGATCATGCTTACGAAACAGAGTACACAGTATGATGATTTTAT 219  
151 CGCTTCTGAGAGCGCTTCCACCCCTGAAGCCCTGCGGCTTCTGGAAGAGTC 210  
220 CGTTTCAATGAAATGTTTCTCAATGAAAGCTAGCAAAATGCTCCAGAAAGAAAGCTT 279  
211 CTAGAGTCTGCAAGGGGGGCTACTACCGCGGGGCAACCTC----- 258  
280 TTAAGACATGAGAGGAGCCCTTGTATCTCAAGAGCCGGAACATCAAGCTGCT 339  
259 ---CACCGCTGAGCGGAGAGCTGAGAGACTTCCCGAGCTTGCAGAGCTTGCAGG 315  
340 AAACAGATCATGTAGAAATTTGATGGGAAATGCGCAACCGCTGAAGAAATTGCTCA 399  
316 CTCTCTGCTCGGGGCTTACACCGCGCGCGGCTGAGCTCAATGCTTGGGGAGGCG 375  
400 TTGAAGGAAATAGAACCATATACAGAGGCGCATTTGCAAGTATCCGATTTGACTTCA 459  
376 GTGCGCGCGGTGAGCGGAGACGTCGGAAGGTCTCTCCGCTCTTCCCGGAAAGC 435  
460 GAACCTGAGTTGAGCGGCAAGTATGAGTATGAGTATGTTTGGCATTGAGGCA 519  
436 CCCAAGAGAA-----GGAGCTTTTGCCTCGCCAGAGGCTCTCCCGCAG 483  
520 GATCTGCAAAAGCTTCTAGCAGGAAATCTTGTATGAAGGATGCGAAATCATGAT 579  
484 GCGCTGAGACCGGGGGTGTGAACCAAGCCCTCATGAGCTCGGGGCAAGCTTGCCTG 543  
580 GAAACGATTCAGGTGAAATTTCAACCAAGCATATGATCTAAGTTCAGCTATCTTACA 639  
544 CCGAAGCGGCGGCTTGCAGGGGCTGCGCCCTAGGGGCTTCTG 587  
640 CCAACTTCTCCAAAGTGTGAGACTTGTCCGATCCAAAGCTTTTG 683

RESULT 15  
US-09-134-000C-1710  
Sequence 1710, Application US/09134000C

Patent No. 6617156  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucelte-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-032  
CURRENT APPLICATION NUMBER: US/09/134,000C  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/055,778  
NUMBER OF SEQ ID NOS: 6812  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1710  
LENGTH: 1188

TYPE: DNA  
ORGANISM: Enterococcus faecalis  
US-09-134-000C-1710

Query Match 6.4%; Score 62.4; DB 4; Length 1188;  
Beef Local Similarity 48.8%; Pred. No. 0.0023;  
Matches 211; Conservative 0; Mismatches 206; Indels 15; Gaps 1;

8 CCGGCGGAAAGCCCTCTGCTGCTGATCCGGAAGAGCCCGCCCTCCCTGCGGG 67  
53 CATTCAGAGAAATTTTATGCTTGTATGAACGAGAAAGCAATTTACCTTGGCGAG 112  
68 GAGAGAGACCTTACCGGCTCGATCTCCAGGCTCTTGTGACAGACCGGGTGG 127  
113 CGAATACAGATGATATGATTTGATTTGATTTCTGAATATGCTACAACTCCGCTAG 172  
128 AGCAGGCTCCCTCTATTAACCGCGCTTGTGAGGCTTCCACCTGAAGGCTGG 187  
173 ATACAGTATGATTTATTTATGATTTATGAAATGTTTCCAGATTCAGATTTAG 232  
188 CCGGCTTCCCTGGAAGGCTTATAGGCTGTGACAGGGGCGGCTACTACCGCGG 247  
233 CGAAGCGCCAGATATTAATTAATTTGAAAGCTTGGAAAGGTTAGTTACTATTACAG 292  
248 C-----GGAACCTTCAACCGCTGCGCGGAGCGTGAAGAGCTTCCC 292  
293 CGGTAATTAAGTGGAGCGCAACAGATTGTTCAAGATTGGTGGAAATGCTG 352  
293 CAGCTTCCGCGAGCTTGGGGGCTTCTGCTGCTGAGCTTACACCGCGCGGCTGG 352  
353 ACACAAATGAAATTTTGGAGTTTAAAGAAATGCGCCCTATACGCTGTGCAATTG 412  
353 CTTCCATGCTTGGGAGCGGCTGAGCGGCTGAGCGGAACTGCGAGGCTCTCT 412  
413 GCAATATGCTTATCTCGGAAACAGCAATTTGACGTTACGTTACGTTAGTGA 472  
413 CCGGCTCTCG 424  
473 GCGTTTATTGG 484

Search completed: January 20, 2005, 04:39:48  
Job time: 134 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2005, 04:17:41 ; Search time 643 Seconds  
(without alignments)  
8712.661 Million cell updates/sec

Title: US-09-938-901A-1

Perfect score: 975

Sequence: 1 atgagggcctgcggcggaagc.....tagtcccccctccgcagcga 975

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	974	99.9	975	US-09-938-901-1	Sequence 1, Appli
2	112.4	11.5	939	US-10-156-761-4691	Sequence 4691, Ap
3	112.4	11.5	9025608	US-10-156-761-1	Sequence 1, Appli
4	104	10.7	1854	US-09-925-301-484	Sequence 484, App
5	104	10.7	1878	US-09-925-301-484	Sequence 1583, Ap
6	92.8	9.5	5850	US-10-437-963-1325	Sequence 1325, Ap
7	91.2	9.4	5850	US-10-437-963-1325	Sequence 1325, Ap
8	89.2	9.1	1660	US-10-425-114-32132	Sequence 32132, A
9	89.2	9.1	1749	US-10-425-114-32132	Sequence 32132, A
10	86.6	8.9	912	US-10-216-817-31	Sequence 31, Appl
11	86.6	8.9	1312	US-10-216-817-21	Sequence 21, Appl
12	86.2	8.8	1425	US-10-437-963-35788	Sequence 35788, A

C 13	82	8.4	9909	8	US-08-961-527-12	Sequence 12, Appl
C 14	82	8.4	9909	16	US-10-158-844-12	Sequence 12, Appl
C 15	81.2	8.3	968	9	US-09-864-865-46	Sequence 46, Appl
C 16	72.6	7.4	2256646	17	US-10-470-565-1	Sequence 1, Appli
C 17	71.4	7.3	930	15	US-10-156-761-4577	Sequence 4577, Ap
C 18	71.4	7.3	9025608	15	US-10-156-761-1	Sequence 1588, Ap
C 19	68.2	7.0	991	16	US-10-398-221-1588	Sequence 3596, Ap
C 20	68.2	7.0	3155	16	US-10-398-221-3596	Sequence 3629, Ap
C 21	66.8	6.9	11058	15	US-10-156-761-3629	Sequence 42, Appl
C 22	66.8	6.9	18876	15	US-10-328-079-42	Sequence 34, Appl
C 23	66.8	6.9	61944	15	US-10-329-079-34	Sequence 56678, A
C 24	66.4	6.8	953	18	US-10-425-115-56678	Sequence 5759, Ap
C 25	65.4	6.7	4667	18	US-10-723-860-5759	Sequence 5404, Ap
C 26	64.6	6.6	1330	15	US-10-156-761-5404	Sequence 1, Appli
C 27	63.8	6.5	8244	15	US-10-402-842-3	Sequence 1, Appli
C 28	63.8	6.5	47988	15	US-10-402-842-1	Sequence 23116, A
C 29	63.6	6.5	2428	17	US-10-437-963-23116	Sequence 17, Appl
C 30	62.8	6.4	2731748	17	US-10-297-465A-1	Sequence 23, Appl
C 31	62.4	6.4	32768	9	US-09-070-927A-17	Sequence 17, Appl
C 32	62.2	6.4	1327	15	US-10-458-201-23	Sequence 1, Appli
C 33	62.2	6.4	15120	15	US-10-458-201-1	Sequence 50882, A
C 34	61.8	6.3	1949	17	US-10-437-963-50882	Sequence 1, Appli
C 35	61.4	6.3	1230025	16	US-10-289-762-1	Sequence 1, Appli
C 36	61.2	6.3	975	10	US-09-938-901-1	Sequence 8597, Ap
C 37	60.6	6.2	1041	17	US-10-437-963-8597	Sequence 12606, A
C 38	60.6	6.2	1484	16	US-10-424-559-105239	Sequence 5, Appl
C 39	60.4	6.2	566	17	US-10-437-963-12606	Sequence 19, Appl
C 40	60.4	6.2	2036	17	US-09-938-901-5	Sequence 1, Appli
C 41	59.8	6.1	1029	10	US-10-205-032-19	Sequence 9041, Ap
C 42	59.8	6.1	5355	15	US-10-205-032-1	Sequence 4667, Ap
C 43	59.4	6.1	60196	15	US-09-864-408A-9041	
C 44	59.4	6.1	272	11	US-10-156-761-4667	
C 45	58.8	6.0	1122	15	US-10-156-761-4667	

## ALIGNMENTS

RESULT 1

US-09-938-901-1

Sequence 1, Application US/09938901

Publication No. US20030008291A1

GENERAL INFORMATION:

APPLICANT: Kuramitsu Seiki.

FILE REFERENCE: PH-1261-US

TITLE OF INVENTION: GENE ENCODING DNA REPAIR ENZYME

CURRENT APPLICATION NUMBER: US/09/938, 901

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: JP2001-47762

PRIOR FILING DATE: 2001-02-23

NUMBER OF SEQ ID NOS: 17

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1

LENGTH: 975

TYPE: DNA

ORGANISM: Thermus thermophilus

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(975)

US-09-938-901-1

Query Match 99.9%; Score 974; DB 10; Length 975;  
Best Local Similarity 100.0%; Pred. No. 3.9e-220;  
Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2	TCGAGGCGTGGGGAAGCCCTTCGCTGTACCGGAAACGCCGCCCTCCCT	61
DB	2	TGAGAGCCCTGGGGAAGCCCTTCGCTGTACCGGAAACGCCGCCCTCCCT	61
QY	62	GGCGGGGGAAGAACCTTACCGCGTCTGTCTCCAGAGTCTTTCGACAGACCC	121
DB	62	GGCGGGGGAAGAACCTTACCGCGTCTGTCTCCAGAGTCTTTCGACAGACCC	121

```

QY      122 GGGTGGAGAGGCGCCCTTCTATTACCGCCGCTTTCTGAGGCGCTTCCCAACCCCTGAAG 181
Db      122 GGGTGGAGAGGCGCCCTTCTATTACCGCCGCTTTCTGAGGCGCTTCCCAACCCCTGAAG 181
QY      182 CCTTGCGCGCGCTTCTTCCCTGAAGAGGTCTTAAAGGCTGAGGAGGGGCGGGCTACTACC 241
Db      182 CCTTGCGCGCGCTTCTTCCCTGAAGAGGTCTTAAAGGCTGAGGAGGGGCGGGCTACTACC 241
QY      242 GCGGCGCGGAGACACCTCCACCGCCCTGCGGCCGAGAGGCTTGAGAGACTTCCCGGAGCTTCC 301
Db      242 GCGGCGCGGAGACACCTCCACCGCCCTGCGGCCGAGAGGCTTGAGAGACTTCCCGGAGCTTCC 301
QY      302 CCGAGCTTCCGAGGCGCTTCTGCTGCTGAGGCTTACACCGGCGGCGGCGGCTTCCATCG 361
Db      302 CCGAGCTTCCGAGGCGCTTCTGCTGCTGAGGCTTACACCGGCGGCGGCGGCTTCCATCG 361
QY      362 CTTTCCGAGGAGCGGGGTGCGCGCGGCTGAGCGGAAAGTCCCGAGAGGCTTCTCCGCGCTG 421
Db      362 CTTTCCGAGGAGCGGGGTGCGCGCGGCTGAGCGGAAAGTCCCGAGAGGCTTCTCCGCGCTG 421
QY      422 TCGCCCGGAGAAAGCCCGAGAGAGAGGCTTTTCGCGCCCTGCGCGCGGCTTCTCCCG 481
Db      422 TCGCCCGGAGAAAGCCCGAGAGAGAGGCTTTTCGCGCCCTGCGCGCGGCTTCTCCCG 481
QY      482 AGGCGCTGAGACCGGCGGCGGTGAGAACAGGCTTCAATGAGCTCGGCGCAAGCTTCC 541
Db      482 AGGCGCTGAGACCGGCGGCGGTGAGAACAGGCTTCAATGAGCTCGGCGCAAGCTTCC 541
QY      542 TCGCGAAACGCGCCGCTTGGGCGGCTGCGCCCTGAGGCGCTTTCGCGGCGGAGAGG 601
Db      542 TCGCGAAACGCGCCGCTTGGGCGGCTGCGCCCTGAGGCGCTTTCGCGGCGGAGAGG 601
QY      602 CCGCCGCGGCGCTACCGCGCGCCAGAGAGCGCGGCGAGAGAGCGCTTCTGCGCG 661
Db      602 CCGCCGCGGCGCTACCGCGCGCCAGAGAGCGCGGCGAGAGAGCGCTTCTGCGCG 661
QY      662 TCGCTCTCTCGGCGGAGAGGCGGTGCACTTGAAAGGCTTGAAGGCGCTTCCAGGCG 721
Db      662 TCGCTCTCTCGGCGGAGAGGCGGTGCACTTGAAAGGCTTGAAGGCGCTTCCAGGCG 721
QY      722 TCTAAGGCGCGCTTCTTCCCTTGAAGGCTTCCCGGCGGAGAGCGGCTTCCGCG 781
Db      722 TCTAAGGCGCGCTTCTTCCCTTGAAGGCTTCCCGGCGGAGAGCGGCTTCCGCG 781
QY      782 TGAAGTCTAAGGCGCTTGAAGGCGGTGCGCAAGCGCTTCAACCGGAGCTTCCGCTG 841
Db      782 TGAAGTCTAAGGCGCTTGAAGGCGGTGCGCAAGCGCTTCAACCGGAGCTTCCGCTG 841
QY      842 AGGTGCGGCGGCGCTTTCGAGAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
Db      842 AGGTGCGGCGGCGCTTTCGAGAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
QY      902 AGCTCATGAGAGAGGTCTCGAGAGGCGCTTCCCTCGCTCATGAGGCGGTAGTCC 961
Db      902 AGCTCATGAGAGAGGTCTCGAGAGGCGCTTCCCTCGCTCATGAGGCGGTAGTCC 961
QY      962 CCTTCCGAGAGCA 975
Db      962 CCTTCCGAGAGCA 975

```

```

RESULT 2
US-10-156-761-4691
; Sequence 4691, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI

```

```

; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4691
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(939)
US-10-156-761-4691

```

```

Query Match      11.5%; Score 112.4; DB 15; Length 939;
Best Local Similarity 52.1%; Pred. No. 2,6e-17;
Matches 369; Conservative 0; Mismatches 306; Indels 33; Gaps 4;

```

```

QY      4 GAGGCTTGGCGGAAAGCCCTCTCGCTGATACCGGAGAAAGCGCCCGCTTCCCTG 63
Db      73 GAGGCTTGGCGGAAAGCCCTCTCGCTGATACCGGAGAAAGCGCCCGCTTCCCTG 63
QY      64 CGGCG---GAGAGAGAGAGCTTACCGGCTGCTGATCCGAGGTCTTTCGAGAGAGC 120
Db      133 CGGCGCGCGAGCGCGGCGCTTGGGCGGTATGCTAGTCAAGTTCATGTTGAGAGAGCA 192
QY      121 CGGCTGAGAGAGCGCTTCCCTATTACCGCGCTTTCGAGAGGCTTCCACCTGAAG 180
Db      193 CGGCTCAACCGCTGATGCTGCTGCTTACAGAGAGTGCCTCCCGTGGCGCCCGCGCG 252
QY      181 GCGCTGCGCGGCGCTTCCCTGAGAGAGGTCTTAAAGGTCTGCGAGGCGGCGCTTACTAC 240
Db      253 GACCTGCGCAAGAGAGGCGCGCGCGAGAGCGGTCCGCGCTGGGCGCGCTCCGCTACCC 312
QY      241 CGGCGGCGGAAACCTTCAACCGCTTGGCC-----GAAAGCTGAGAG 285
Db      313 CGGCGCGCGCTGCGCTGAGAGCGCGCGCGCTGCGCTTACAGAAAGCGCAACGCGAGC 372
QY      286 CTTCCCGGAGCTTTCGCGAGCTTGGGCGCTTCTGCTGAGGCTTATACCGCGGCG 345
Db      373 GTACCAAGAGAGAGCGCGAGCTGCTGCGCGCTGCGCGCGAGATCGGCAATACCGCGCG 432
QY      346 GCGGTGCTCTCATCGCTTTCGAGAGCGGAGGCGGCGGTGAGAGAGAGAGAGAGAGAG 405
Db      433 GCGGTGCGGTGCTGCTGAGAGCGCGAGCGCGCTTCTGAGAGAGAGAGAGAGAGAG 492
QY      406 GTCTCTCCCGGCTTTCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
Db      493 GTCTTTCGCGCGCGCTTCAACCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 552
QY      460 -----CTGCGCAAGAGCTCTTCCCGAGGCGGTGAGAGAGAGAGAGAGAGAGAG 510
Db      553 CGCAAGCTGCGCGCGCGCTGCTTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 612
QY      511 GCGCTCATGAGAGCTTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570
Db      613 GCGCTCATGAGAGCTTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672
QY      571 CCGCTAAGGAGCTTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 630
Db      673 CGATGCGCGGCGAGTGCCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 732
QY      631 CGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 678
Db      733 CGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780

```

RESULT 3



Db 1034 GAGCCTGTGCCGGG 1047

```

RESULT 5
US-09-925-301-484
Sequence 484, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OR INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ. ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 484
LENGTH: 1878
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1446)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-484

```

Query March	Score 104;	DB 9;	Length 1878;
Best local similarity	10.7%;		
	Pred. No. 2 2e-15;		
	54.0%;	Indels 30;	Gaps 3

QY	GGAGAGAGACCTTACCGCGCTCTGTCTCCAGAGTCTTCTGACACAAACCGGAGTGA	128
Db	GGACAGCGGGCATATGCTGTGTGGGCTTCAGAGGTATGCTGACACAAACCGAGTTGC	568
QY	129 GCAGGCGCTTCCCTCTATTACCGCGCTTCTAGAGCGCTTTCCACCTGAAGCGCTGGC	188
Db	CACTGTGATCTACTATTATCCGATGAGATGCAGAAAGTGGCTCACTGCAGGACCTTGGC	628
QY	189 CGCGGCTTCCCTGGAAGAGTCTTAGGATCTGCGAGGGGCGGGCTTACTACCGGCGGC	248
Db	629 CAGTGTCTCCCTGGAAGAGTGAATCACTCTGGGCTGGCTGGGCTACTATTCTGTGG	688
QY	249 GGAACACTTCCACCGCGCTGGCCCGAAGCGTGGAGAGCTTCCCCGAGCTTC-----	300
Db	689 CCGCGGCTCAGAGAGGAGCTGGAAAGTGTGAAGAGTCAAGGCGGCCACATGCCACG	748
QY	301 -----GCCGAGCTTTCGGGGGCTTCTGTCTCGGGCTTACACCGGGGGGGGAGT	350
Db	749 TACACAGAGACCTTGCAGACGCTCCGCTGCGGGCGGGGCCCTTACACAGCTGGAGGCAT	808
QY	351 GGCCTTCATGCGCTTTCGGGAGCGGGGTGGCGCGGTGGAGCGGGAACGTCGGAAGTCT	410
Db	809 TGCCCTTATCCGCTTTTGGCCAGGCAACCGGTGTGTGGATGCGAACGTGAGCAGGGGTCT	868
QY	411 CTCGCCGCTCTTGGCCC--GGAAAGGCCCAAGAGAGAGACTTTTGGCCCTGGCCAG	468
Db	869 GTGGCGGTGTCGAGGCCATTGGTGTGTATTCACAGACAGACCTGTGTTTCCAGACGCTTG	928
QY	469 GGCCTCTTCCCGAGGGCGTGAACCG-----GGGTGTGGAACAGAGCGCTCAT	518
Db	929 GGGTCTAGCCAGCAGCTGTGTGGACCCAGGCCCGGCGAGGATTTCAACCAAGACAGCAT	988
QY	519 GGAAGCTCGGGGCAACGCTCTGCTGCGGAACGAGCGCCCGTGTGCGGGGCTGCCCCCTAG	578
Db	989 GGAAGCTAGGGGCAACGCTGTGTATCCCAAGCGGCCCACTGTGACGCAAGGCGCTGTGGA	1048
QY	579 GGCCTTCTGCGGAG 592	
Db	1049 GAGCTGTGCGGG 1062	

RESULT 6  
US-09-974-300-1583  
Sequence 1583, Application US/09974300

Query Match	9.5%	Score 92.8;	DB 9;	Length 528;
Best Local Similarity	53.4%;	Pred. No. 1.2e-12;		
Matches 227;	Conservative	0;	Mismatches 183;	Indels 15;
				Gaps 1

43	AAAGATTAA	TTTCTTG	TGATG	AGCAAGAAAA	CGGGATCTG	CTTGGCGCT	CAGACGCG	102
Dd								
Qy	77	ACCTTAA	CGCGCTC	TGCTGCT	CCGAGTCTCT	GTGAGACAG	ACCGGGTGA	GCAGCGCC 136
Dd	103	ACCTTTAA	GAGTGTG	GGGTGTG	CGAAGTAA	TGCTTCAG	CAGACGAGAGT	GGACCGGTGA 162
Qy	137	TCCCTTA	TTAACCCG	CGTTTCT	GTGAGCGCTTT	CCCACTCT	GAGGCGCTTG	GC CGCGGCTT 196
Dd	163	TTCTTCT	CTTCAACAA	CTTTATG	CAAAAAGTTTC	CGACCGTGA	AGGCGCTCG	CTGTAAGCGG 222
Qy	197	CCCTGA	AAGAGTCT	TTAGGCT	GTGCAGGGCGG	CGGCTACTTA	CCGCGGGGCG	AGAAACAC 256
Dd	223	ACGAAG	AAAAAGT	GTCTTAA	GCGCTGGAA	AGGAGCTGG	AGATCTTA	TTCAAGGCTTCGGAATC 282
Qy	257	TCACACG	-----	CTGGCC	CGAAGCGTGA	GAGACTTCC	CCCGAGCTTG	301
Dd	283	TGCNAG	AGGCTGTG	CAGGAA	ATTACAA	AGATACG	AGNGTGTG	TCCCTTTCAAAAG 342
Qy	302	CCGAGCTT	CGGGGGCTT	CTGCTCT	CGGAGCTT	CACACGCG	CGGCGGTG	AGCTTCATCG 361
Dd	343	AGGAAT	TGCGGAC	GTTAAAG	AGCGCTCGG	CCCTTTA	CAAGAGGCG	CGGTCTGACATCG 402
Qy	362	CGTTG	GGGAGCG	GGGTGCG	CGGTGA	CGGAAAGT	TCGAGGGT	CTCTCCCGGCTCT 421
Dd	403	CTTACAT	ATAGCGG	GTCCCG	CGGTGA	CGGAAATG	TCATGCG	GGGTATGTACGAGATTC 462
Qy	422	TCGCC	426					
Dd	463	TTTCC	467					

RESULT 7  
US-10-437-963-1325  
; Sequence 1325, Application US/10437963  
; Publication No. US20040123342A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.

```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 1325
; LENGTH: 5850
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_101202C.1
US-10-437-963-1325

```

```

Query Match          9.4%; Score 91.2; DB 17; Length 5850;
Best Local Similarity 54.4%; Pred. No. 1.9e-12;
Matches 211; Conservative 0; Mismatches 168; Indels 9; Gaps 1;

```

```

QY 65 GGGGGGAGAGACCTTACCGGCTGCTCCGAGGCTCTTCGACAGACCCGGG 124
DB 263 GGAAGAGAGAGGCGGTACGCGGTGTGGGTGCGAGGTGATGTCGACGACATACCGGG 322
QY 125 TGGACAGAGCCCTCCCTTATTAACCGCGCTTCTGAGCGCTTCCACCGTAAAGCC 184
DB 323 TGCCGCTGTGTGCTGACCTACTACTCCGCTGGATGAGCCCGCGCCACCGTGAAGCC 382
QY 185 TGGCGCGGCTTCCCTGGAAGAGTCTTAAGGCTGCGAGGGGCGGCTACTACCGGC 244
DB 383 TCGCGCGTCCGACGAGGAGGTGAACGAGATGTGGCGGCTCGGCTATTAACCGGA 442
QY 245 GGGGGGAGACCTTCAACCGCTGCGCCGAGAGCT-----GAGAGAGCTTCCCGCA 295
DB 443 GGGCTCGATTCTTCTCGAGGAGCAAGCAAAATGTTGAAAAAGGCGAGTCCCTCGCA 502
QY 296 GCTTCCGAGCTTGGGGGCTTCTGCTGCTCGGCTTACACCGCGCGCGGTGGCT 355
DB 503 CAGCATCAGCACTCCGTGAAGTTCGTGGCATTTGGGATTAACAGAGAGACCATTTCT 562
QY 356 CCATGCTTGGGGAGCGGGTGGCGCGGTGACCGGAACCTCCGAGGCTTCTTCC 415
DB 563 CCATGCTTCAATGAGGTGTCTCTGTGTGATGGAATGCGAGTTATCAGAGGC 622
QY 416 GCCTTTGCGCGGAGAGCCCAAGGA 443
DB 623 TTACGCTATTCTCTGATTAACCAAGA 650

```

## RESULT 8

```

US-10-425-114-32132
; Sequence 32132, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 32132
; LENGTH: 1660

```

```

; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73265D07_FLI
US-10-425-114-32132

```

```

Query Match          9.1%; Score 89.2; DB 16; Length 1660;
Best Local Similarity 50.7%; Pred. No. 6.9e-12;
Matches 282; Conservative 0; Mismatches 253; Indels 21; Gaps 2;

```

```

QY 60 CTGGCGGGGAGAGAGACCTTACCGCGCTGCTCCGAGTCTCCGAGTCTTTCGAGAGAC 119
DB 278 CGGACGCGAGAGAGAGAGCGGTACCGCGGTGTGGGTGTCCGAGTATCTGACGAGAC 337
QY 120 CCGGTGAGACAGAGCCCTCCCTATTAACCGCGCTTTCGAGCGCTTTCACCTGAA 179
DB 338 GCGGGTCCCGTGTGTGTGCTTACTACAGCGGTGATGGCGCGGTGCGACCGTACG 397
QY 180 GGCCTTGGCGCGGCTTCCCTGGAAGAGTCTTAGAGTCTGGCAAGGGGCGGCTACTA 239
DB 398 AAGCTTGGCGCGGCTTACCGAGAGAGGTGAACGATGTGGCGGCTTGGCTACTA 457
QY 240 CCGCGCGCGGAGAACCTTC-----ACCGCTGCGCGGAGAGCGTGAAGAGCTTCC 290
DB 458 CCGTAGGCTCGATTCTGCTGAGAGAGCAAGACGATCATCGAAGAGGGGTGTTCC 517
QY 291 CCGAGCTTCCCGAGCTTGGGGGCTTCTGTGCTCGGCTTACACCGCGCGCGT 350
DB 518 TTGACAGGCGGTAAAGCTTCTGAGAGGTGTGTGCAATTGGGATTTACACAGCTGAGCAAT 577
QY 351 GGCCTTCATCGCTTCCGAGGAGCGGGTGGCGCGGTGAGCGGAGACGTCCGAGGGTCT 410
DB 578 CGCTTCAATGCTTCAAGAGAGTGTCCAGTTGTGATGGAATGTAATACAGATCAT 637
QY 411 CTCGCGCTTCTTCCCGGAGAGCCCAAGAGA-----GAGCTTTTCG 458
DB 638 CAGCAGGCTTTACACATTTGTGACAAACCAAGAAATCTCAACGTAAGAGATTCG 697
QY 459 CCTCGCCAGAGGCTCTCTCCGAGGCGGTGACCTCGGGGCTGTGAACCAAGCCCTCAT 518
DB 698 GGAACCTGTGAGGTCAATGTGTGACCTTTGAGACCGAGAGACTTCAACCAAGCGATGAT 757
QY 519 GGAAGTCCGGGCGAGGCTGTGCTGCGCAAGCGCCGCTTGGCGGGCGCTCCAGG 578
DB 758 GAGGTTAGAGCAACTTATTTGTAAGCAAGAAAGCTGTGTCTCAATGCCAGTCTC 817
QY 579 GGCCTTCTGCGCGGG 594
DB 818 TAGCCACTGCCAAGCG 833

```

## RESULT 9

```

US-10-425-115-92673
; Sequence 92673, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 92673
; LENGTH: 1749
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_184509C.1
US-10-425-115-92673

```

Query Match 9.1%; Score 89.2; DB 18; Length 1749;  
 Best Local Similarity 50.7%; Pred. No. 6.9e-12;  
 Matches 282; Conservative 0; Mismatches 253; Indels 21; Gaps 2;

```

QY 60 CTGGCGGGGGAAGAACCCCTTACCGGCTCTGCTCCGAGTCTCTTCTGACAGAC 119
DB 278 CGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 337
QY 120 CCGGCTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 179
DB 338 GCGGGTCCCGTGGTCTGCTTACTACGAGCGGTGAGAGAGAGAGAGAGAGAGAG 397
QY 180 GGGCTTGGCGGCTTCCCTGAGAGAGTCTTGAAGGCTGTCGACAGAGAGAGAGAG 239
DB 398 AAGCTGCGCGCGCTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 457
QY 240 CCGGCGGGGGAACACCTCC-----ACCGCTGGCCCGAAGCGGTGAGAGAGAGAG 290
DB 458 CCGTACGGCTGATTTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 517
QY 291 CCGGAGCTTCCCGAGCTTGGGGGGCTTCTGCTGCTGAGAGAGAGAGAGAGAGAG 350
DB 518 TTGACAGCGCTTACCGCTTCTGAGAGTCTGAGATTTGAGAGAGAGAGAGAGAG 577
QY 351 GGGCTTCCATCGCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 410
DB 578 CGCTTCAATTGCTTCAACGAGAGTGTCCAGTGTGGATGGAATGTGATACAGAGCAT 637
QY 411 CTCGCCCTCTTGGCGGGGAAAGCCCAAGAGAA-----GGAGCTTTTGGC 458
DB 638 CAGCAGAGGCTTTACACCATTTGTCGACACCAAGAGATCTTCAACAGTGAAGAGAT 697
QY 459 CCTGCGCCAGGAGCTCTCCCGAGAGGCTGAGACCGGGGGTGTGAAACAGAGCCCTCAT 518
DB 698 GGAACCTCTGAGTCAAAATGTGTTGACCTTTTGAACAGAGAGATTTCAACAGAGAGAT 757
QY 519 GAGAGCTGGGGGCAAGGTCTGCTGCGCCGAAAGCGCCCGTGGGGGCTGCCCTTAG 578
DB 758 GAGGTAGAGAGCACTTATGTAGCAAGCAAAAGCTGTGCTCTCAATGCCAGTCTC 817
QY 579 GGGCTTCTGCGGGGG 594
DB 818 TAGCAGCTGCAAGCG 833

```

RESULT 10  
 US-10-216-817-31

```

; Sequence 31, Application US/10216817
; Publication No. US20030129619A1
; GENERAL INFORMATION:
; APPLICANT: GICOUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; TITLE OF INVENTION: IN GENES OF THE MULT FAMILY
; FILE REFERENCE: 03495.0233-00000
; CURRENT APPLICATION NUMBER: US/10/216,817
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 912
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-216-817-31

```

Query Match 8.9%; Score 86.6; DB 15; Length 912;  
 Best Local Similarity 50.2%; Pred. No. 3.1e-11;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

```

QY 22 CTCCTGCTGTAACGGGAAACCGCCCGCCCTCCCTGCGGGG---GAGAGAGAC 78
DB 61 CTCTCTGCTTGTATAGGATGCGACCGGATCTGCTTGGCGAGAGACCGCGTGTACG 120
QY 79 CTTTACCGGCTCTGCTTCCGAGGTCTTCTTGCACAGACCCGGGTGAGACAGGCCCC 138
DB 121 CCGTGGCAATCTTGTTCAGAGGTTCAATGTGACAGACACCGCGCCCGCGGTGCTG 180
QY 139 CCTTATACCGCGCTTTCTGAGAGGCTTTCCACCCCTTAAAGCCCTGAGCGGCTTTC 198
DB 181 GCGATGTGCGGAGCTGGGTGCGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 199 CTGGAAGAGTCTTGAAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 258
DB 241 ACCGCGATGTGTTATCGCGCTGGGCAAGCTGAGCTATCCAGAGCGACCAAGGCTTA 300
QY 259 CACGCGCTGCGCGAAGCGTG-----GAGAGAGCTTCCCGAGCTTGGC 303
DB 301 CACGAGTGGCGCACCGTATCGCGCGACACACATGATGTGCTCCAGATATCGAG 360
QY 304 GAGCTTGGGGGCTTCTGAGTCTGAGCTTACACCGCGAGCGGTGAGCTTCAATGCC 363
DB 361 ATCTGTACCTTGGCGCGGTGAGAGTACACCGCGCGGTGAGGTGTGCTTCCGT 420
QY 364 TTGAGAGAGCGGTGAGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422
DB 421 TACGCGCAGCGGTGCGGTGTGAGACCAATGTGCGGTGAGGTGCGGTGCGGT 480
QY 423 -----CGCCCGGAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471
DB 481 CAGGCGCGCGCGAGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 472 CTCCTCCCGAGAGGAGTGAACCCCGGGGTGTGAAACAGAGCCCTCAAGAGAGAG 531
DB 541 CTGTTGCGCACCGGAGAGAGCGCGCTGAATTTTGGTGGCTGATGAGATGGGTGCG 600
QY 532 AGGCTGTGCTGCGGAAAGCGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 591
DB 601 ACGGTGTGACCGCGCGCACCGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 592 GGAAGAGAGAGCGCGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651
DB 661 CCGCATGCGCGGTATTCGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717
QY 652 CTCGTGCGCTGCTCTCTGCGGGCG 678
DB 718 GGAACCGAGCGCAAGTCCGCGAGCG 744

```

RESULT 11  
 US-10-216-817-21

```

; Sequence 21, Application US/10216817
; Publication No. US20030129619A1
; GENERAL INFORMATION:
; APPLICANT: GICOUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; TITLE OF INVENTION: IN GENES OF THE MULT FAMILY
; FILE REFERENCE: 03495.0233-00000
; CURRENT APPLICATION NUMBER: US/10/216,817
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1312
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-216-817-21

```

Query Match 8.9%; Score 86.6; DB 15; Length 1312;  
 Best Local Similarity 50.2%; Pred. No. 2,9e-11;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

```

QY 22 CTCCTCGCTGTGTACCGGAAAAAGCCCGCCCTCCCTCGCGGGG---GAGAAGAC 78
DB 261 CTTCTCGCTGTGTATCAGGATGCAACCGGATCTCCCTGGGAGAGCCGGTGTACG 320
QY 79 CTTTACCGGCTCTGTGTCTCCAGAGTCTTTGACAGACACCGGGTGAAGACCCCTC 138
DB 321 CGGTGAGAGATCTGTGTACGAGTTCATGTGACAGACACCGCGCCCGCGGTGTG 380
QY 139 CCTATTACCGCGCTTTGTGAGAGCTTTCCACCCCTAAGAGCCCTGCGCGGTTTC 198
DB 381 GGAATGTGCGGAGCTGGGTGCGCGGTCGCGCCACCGCTCGACCCGCAAGCGCAGC 440
QY 199 CTGGAAGAGTCTTATGAGTCTTGCAAGGAGCGGCTACTACCGCGGCGGACACCTC 258
DB 441 ACCGCGGATGTGTACGCGCTGTGGGCAAGCTGGGTATCCAGGAGACCAAGCCTTA 500
QY 259 CACCGCGTGGCCCGAAGCGTG-----GAGAGCTTCCCGGAGCTTTCGCC 303
DB 501 CACGAGTGGCGCACCGTCACTCCCGCGACCAATGATGATGTCGCGACGATATCGAG 560
QY 304 GAGCTTGGGGGCTTCTGTGTCTCGGCGCTTACACCGCGCGCGGTGGCTTCATCGCC 363
DB 561 ATCTGTGTACCTCTGCGCGCGCTGTGGAGCTACACCGCGCGGTGGCGGTTCGCT 620
QY 364 TTGCGGAGACCGGATGCGCGGTGAGACGGGAAAGTCCGAGAGGCTCTTCCGCGCTT- 422
DB 621 TACCGCGAGCGGCTGTGTGTGACACCAATGTGCGCGGTGTGGCTCGCGCGCTT 680
QY 423 -----CGCCGCGGAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471
DB 681 CACGCGCGCGCGACGCGCGGTGCGCATGCGGTGCGCGACCAACGCGCATGCTTGGCG 740
QY 472 CTCCTCCCGGAGGCGGTGACCCCGGGGTGTGAAACCAAGCCCTCATGAGCTCGGGGCC 531
DB 741 CTGTTCGCGACCGGAGAGCGCGCTGAAATTTTGTGCGCTGATGAGTGGGTGG 800
QY 532 ACGGTCTGCTGCTCGGAAAGCGCGCTTGGGAGCGCTGCGCTTAGGGGCTTCTGCGCG 591
DB 801 ACGGTGTGACCGCGCGCGACACCGCGGTGTGCGCTGTGACTGTGTGCGCATGG 860
QY 552 GGAAGAGAGCGCCCGCGCGCTAACCCGCGCGCAAGAGCGCGCGGAGAGAGAGCGC 651
DB 861 CGGCAATGCGGTTATCCGCGGTGCGAGCGGTCC---GCCGCGCGGGGAGAGGCTTACAC 917
QY 652 CTCGTGCGCTGTCTCTCTCGGGCGG 678
DB 918 GGAACCGAGCGCGCAAGTCCGCGAGCG 944

```

RESULT 12  
 US-10-437-963-35788  
 ; Sequence 35788, Application US/10437963  
 ; Publication No. US2004012343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Bouharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 35788  
 ; LENGTH: 1425  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_39677C.1  
 US-10-437-963-35788

Query Match 8.8%; Score 86.2; DB 17; Length 1425;  
 Best Local Similarity 55.7%; Pred. No. 3,6e-11;  
 Matches 191; Conservative 0; Mismatches 143; Indels 9; Gaps 1;

```

QY 65 GGGGAGAGAGAGACCTTACCGCGCTTCTGTCTCCAGGTCTTTGACAGACAGCCGG 124
DB 266 GCGAGAGAGAGAGGCGTACCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 325
QY 125 TGAGAGAGCCCTCCCTTATTAACCGCGCTTTCTGAGGCTTTCCACCTGAAGAGCC 184
DB 326 TACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 385
QY 185 TGGCGCGGCTTCTCTGAGAGAGTCTTATGAGGTCTGAGAGGCGCGGCTACTACCGC 244
DB 386 TCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 445
QY 245 GGGCGGAAACCTCCACCGCGCTGCGCGGAGAGGT-----GAGAGAGCTTCCCGCA 295
DB 446 GGGCTGTGATTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 505
QY 296 GCTTCCGCGAGCTTGGGCGCTTCTGTCTGCGGCTTACACCGCGCGCGGTGAGCT 355
DB 506 CAGCATCAACACTCGGTGAAGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 565
QY 356 CCATGCGCTTGGGAGAGCGGTGTGCGCGGTGTGAGCGGAGAGAGAGAGAGAGAG 398
DB 566 CCATGAGCTTCAATGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 608

```

RESULT 13  
 US-08-961-527-12/c  
 ; Sequence 12, Application US/08961527  
 ; Publication No. US2002003232A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunch  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 391  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/961,527  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brookes, A. Anders  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB340P1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:

Tue Jan 25 11:49:52 2005

us-09-938-901a-1.tmpb

LENGTH: 9909 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-12

Query Match 8.4%; Score 82; DB 8; Length 9909;  
Best Local Similarity 52.2%; Pred. No. 2.6e-10;  
Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

22 CTCCTGCGCTGATCCGGGAAACGCCGCCCTCCCTCGCGGGGGAGAGACCT 81  
5808 CTTCTGCTGGTATGATGAAACAAAGATTGCTTGGAGGAGAAATTAATCT 5749  
82 TACCGCGCTGCTGCTCCGAGGCTCTTCTGACAGACGCCGGGTGAGAGCCCTCCC 141  
5748 TATCATCTGGGTATCGAATATCATGCTTCAGACAGACAGGGGTGATACAGTTATCCCT 5689  
142 TATTACCGCGCTTCTGAGAGCGCTTTCACACCTGAAAGCCCTGGCGCGGCTTCCTG 201  
5688 TACTAGAAAGATTTTGGACTGGTTTCACTGTCGAAAGTCTGGCACTGCGCTGAG 5629  
202 GAAGAGTCTTAGGGTCTGCGAGGGGGGGGCTA-----CTACCGCGG 246  
5628 GAGAGTTACTGAAACCTTGGAGGCTTGGGCTTATTCTCGAGTTCCGAATATGACAG 5569  
247 GCGGAACACCTCCACCGCCTGCGCCGAGACGTGAGAGAGCTTCCCGAGCTTCGCCGAG 306  
5668 GCTGACGCCAGACAGATTAATGACTGACTTGGTGGCCAAATTCGAATACCTATGAGAGA 5509  
307 CTTGGGGGCTTCTGCTGCTCGGGCCTTACACCGCGCGCGGCGCTTCATGCGCTTC 366  
5508 ATTCCAGCTTGAAGAGGATTTGACCTTACACAGAGAGACCATTTCCAGATATGCTTTT 5449  
367 GGGGAGCGGTGCGCGCGGTGACGCGAGACGTCCGAGGGTCTCTCCGCGCTTCG 424  
5448 AACTTGCTGAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 5391

RESULT 14

US-10-158-844-12/c  
Sequence 12, Application US/10158844  
Publication No. US20040029118A1

GENERAL INFORMATION:

APPLICANT: Kunsch et al.  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-R  
COMPUTER: Dell Latitude Pentium 3  
OPERATING SYSTEM: Windows 98  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/158,844  
FILING DATE: 03-Jun-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/961,527  
FILING DATE: 1997-10-30  
APPLICATION NUMBER: US 60/029,960  
FILING DATE: 1996-10-31  
ATTORNEY/AGENT INFORMATION:  
NAME: Hyman, Mark J.  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB340P1D1

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:  
LENGTH: 9909 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-10-158-844-12

Query Match 8.4%; Score 82; DB 16; Length 9909;  
Best Local Similarity 52.2%; Pred. No. 2.6e-10;  
Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

22 CTCCTGCGCTGATCCGGGAAACGCCGCCCTCCCTCGCGGGGGAGAGACCT 81  
5808 CTTCTGCTGGTATGATGAAACAAAGATTGCTTGGAGGAGAAATTAATCT 5749  
82 TACCGCGCTGCTGCTCCGAGGCTCTTCTGACAGACGCCGGGTGAGAGCCCTCCC 141  
5748 TATCATCTGGGTATCGAATATCATGCTTCAGACAGACAGGGGTGATACAGTTATCCCT 5689  
142 TATTACCGCGCTTCTGAGAGCGCTTTCACACCTGAAAGCCCTGGCGCGGCTTCCTG 201  
5688 TACTAGAAAGATTTTGGACTGGTTTCACTGTCGAAAGTCTGGCACTGCGCTGAG 5629  
202 GAAGAGTCTTAGGGTCTGCGAGGGGGGGGCTA-----CTACCGCGG 246  
5628 GAGAGTTACTGAAACCTTGGAGGCTTGGGCTTATTCTCGAGTTCCGAATATGACAG 5569  
247 GCGGAACACCTCCACCGCCTGCGCCGAGACGTGAGAGAGCTTCCCGAGCTTCGCCGAG 306  
5668 GCTGACGCCAGACAGATTAATGACTGACTTGGTGGCCAAATTCGAATACCTATGAGAGA 5509  
307 CTTGGGGGCTTCTGCTGCTCGGGCCTTACACCGCGCGCGGCGCTTCATGCGCTTC 366  
5508 ATTCCAGCTTGAAGAGGATTTGACCTTACACAGAGAGACCATTTCCAGATATGCTTTT 5449  
367 GGGGAGCGGTGCGCGCGGTGACGCGAGACGTCCGAGGGTCTCTCCGCGCTTCG 424  
5448 AACTTGCTGAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 5391

RESULT 15

US-09-864-866-46  
Sequence 46, Application US/09864866  
Patent No. US20020127656A1

GENERAL INFORMATION:

APPLICANT: Lloyd R. Stephen  
APPLICANT: McCullough, Amanda K.  
TITLE OF INVENTION: DNA REPAIR POLYPEPTIDES AND METHODS OF USE  
FILE REFERENCE: 265, 001/0101  
CURRENT APPLICATION NUMBER: US/09/864,866  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/206,279  
PRIOR FILING DATE: 2000-05-23  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO: 46

COMPUTER READABLE FORM:

MEDIUM TYPE: DNA  
COMPUTER: Micrococcus luteus  
OPERATING SYSTEM: Windows 98  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/864,866  
FILING DATE: 2001-05-23  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/961,527  
FILING DATE: 1997-10-30  
APPLICATION NUMBER: US 60/029,960  
FILING DATE: 1996-10-31  
ATTORNEY/AGENT INFORMATION:  
NAME: Hyman, Mark J.  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB340P1D1

INFORMATION FOR SEQ ID NO: 12:



```
Db 260 TGCTGTCCGCCAGACCAACCGAGTGGGGGTCAACGACGACCGCGGCGCTGTTGCGCC 319
QY 164 GCTTTCCACCCCTGAAGGCCCTTGCCCGCGGCTTCCCTGGAAGAGTCTTAGAGTCTGCG 223
Db 320 GCTTCCCGAGTGCCTCAGCATGAGCCGCGGCAACGAGCCGAGCTGCAAGAGCTCGTGC 379
QY 224 AGGGGCGGGCTACTACCGG-----CGGGCGGAACACTTCACCGCC 265
Db 380 GCTCCACGGGGTCTTACCGGAACAAGGCTCCCGGATCTGCGGCTGTCCAGAGCTCG 439
QY 266 TGGCCCCGAAGCCGTGAAGAGCTTCCCGGAGCTTCCCGGAGCTTCCGGGGCTTCTGTGC 325
Db 440 TGGGCGGCAAGACGAGGAGAGTCCCGGCGGCTCGAGGACCTGTGGCGCTGCGCGGG 499
QY 326 TCGGGCTTACACCGGCGGCGGCGGCTTCATCGGCTTCCGGGAGACGGGTGCGCGG 385
Db 500 TGGGCGGCAAGACCGGCTTCTGTGTCTCGGCAACGCTTCCGGCCAGCCCGGATCACCG 559
QY 386 TGGACGGGAACGTCCGGAAGGTCCTCTCCGCTCTTCCGCGGGAAGCCCAAGAGA 445
Db 560 TGGACACGCACTTCCGCGGCTGCGCGGCGCTTGGGTTTCAAGAGAGACCGACCGG 619
QY 446 AGGAGCTTTGCGCTCGGCCAGGAGCTCTCCCGAGGAGCTGGAACCCGAGGCTGGA 505
Db 620 GTAAAGTGAAGCAGCGCGGCTGTGCTTCCCGCGCGGGAAGTGAAGTGTCT 679
QY 506 ACCAGGCTTCAAGAGCTTGGGAGCAAGGTCTGCTGCGGAAACGCGCCGTTGCGGG 565
Db 680 CCCACCGGCTGATCTTCCAGGCGCGCGCTGTGCAAGCGCGCGCGCGCTGCGGCG 739
QY 566 CTTGCCCTTAAGGGGCTTCTGCGGGGGAAGAGGCCCCGGGCGCTACCCGCGCCA 625
Db 740 GGTGCCGATGCGCGCTGTGCTGCTCTTACGCGCGGGGAGACCGACCGAGCGG 799
QY 626 GGAAGCGCGGCGGAAGAGAGCGCTCTGCGCC 661
Db 800 CGCGCGCTGCTGCGCTACGAGCTCAAGCCGCGC 835
```

Search completed: January 20, 2005, 06:10:40  
Job time : 678 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 20, 2005, 04:37:51 ! Search time 26 Seconds  
(without alignments)  
828.975 Million cell updates/sec

Title: US-09-938-901a-2  
Perfect score: 1701  
Sequence: 1 MEAMRKALLAWYRENNARPLP.....VLKRLPLLAHGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/6C\_PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfile01.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	505	29.7	403	4 US-09-107-532A-7308	Sequence 7308, Ap
2	485	28.5	395	4 US-09-134-000C-5115	Sequence 5115, Ap
3	481.5	28.3	535	3 US-08-813-574-2	Sequence 2, Appl
4	477.5	28.1	470	4 US-09-252-991A-23310	Sequence 23310, A
5	477	28.0	375	4 US-09-198-452A-421	Sequence 421, App
6	471.5	27.7	350	3 US-08-813-574-9	Sequence 9, Appl
7	471.5	27.7	350	3 US-09-651-656-37	Sequence 37, Appl
8	471.5	27.7	350	3 US-09-650-855-37	Sequence 37, Appl
9	458.5	27.0	391	4 US-09-583-110-4583	Sequence 4583, Ap
10	457.5	26.9	516	4 US-09-489-039A-11959	Sequence 11959, A
11	450	26.5	347	4 US-09-543-681A-5315	Sequence 5315, Ap
12	443	26.0	360	3 US-09-134-001C-4684	Sequence 4684, Ap
13	442	26.0	346	3 US-09-128-352-4982	Sequence 4982, Ap
14	389.5	22.9	188	4 US-09-629-222A-36	Sequence 36, Appl
15	355	20.9	414	4 US-09-540-236-2578	Sequence 2578, Ap
16	274.5	16.1	221	1 US-08-663-023-17	Sequence 17, Appl
17	274.5	16.1	221	4 US-09-402-959A-10	Sequence 10, Appl
18	274.5	16.1	225	4 US-10-037-927B-14	Sequence 14, Appl
19	240.5	14.1	188	4 US-09-629-222A-34	Sequence 34, Appl
20	199	11.7	188	4 US-09-864-866-43	Sequence 43, Appl
21	191	11.2	185	4 US-09-629-222A-35	Sequence 35, Appl
22	188	11.1	227	4 US-09-107-532A-3843	Sequence 3843, Ap
23	176	10.3	207	2 US-08-808-550-36	Sequence 36, Appl
24	176	10.3	259	2 US-08-808-550-39	Sequence 39, Appl
25	172	10.1	209	2 US-08-808-550-31	Sequence 31, Appl
26	171	10.1	204	2 US-08-808-550-32	Sequence 32, Appl
27	164.5	9.7	211	2 US-08-808-550-37	Sequence 37, Appl

28	164.5	9.7	297	2 US-08-808-550-42	Sequence 42, Appl
29	164.5	9.7	304	2 US-08-808-550-2	Sequence 2, Appl
30	164	9.6	210	4 US-09-710-279-972	Sequence 972, Appl
31	164	9.6	224	4 US-09-710-279-2684	Sequence 2684, Ap
32	160	9.4	224	3 US-09-134-001C-3225	Sequence 3225, Ap
33	159	9.3	209	4 US-09-583-110-4518	Sequence 4518, Ap
34	158	9.3	185	4 US-09-134-000C-3440	Sequence 3440, Ap
35	154	9.1	213	2 US-08-808-550-35	Sequence 35, Appl
36	145	8.5	207	2 US-08-808-550-29	Sequence 29, Appl
37	145	8.5	211	2 US-08-808-550-38	Sequence 38, Appl
38	145	8.5	211	3 US-09-651-656-25	Sequence 25, Appl
39	145	8.5	211	3 US-09-650-855-25	Sequence 25, Appl
40	145	8.5	211	4 US-09-492-709A-296	Sequence 296, Appl
41	142	8.3	216	4 US-09-489-039A-9560	Sequence 9560, Ap
42	137.5	8.1	184	4 US-09-629-222A-33	Sequence 33, Appl
43	134	7.9	207	2 US-08-808-550-30	Sequence 30, Appl
44	128	7.5	628	4 US-09-252-991A-30904	Sequence 30904, A
45	127.5	7.5	212	4 US-09-252-991A-17572	Sequence 17572, A

## ALIGNMENTS

RESULT 1  
US-09-107-532A-7308  
Sequence 7308, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
City: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107, 532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085, 598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-8277  
TELEFAX: (781)893-5007  
INFORMATION FOR SEQ ID NO: 7308:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...403  
SEQUENCE DESCRIPTION: SEQ ID NO: 7308:  
US-09-107-532A-7308

Query Match 29.7%; Score 505; DB 4; Length 403;  
Best Local Similarity 32.7%; Pred. No. 2,3e-42;  
Matches 118; Conservative 66; Mismatches 113; Indels 64; Gaps 11;

QY 9 LAWRRENARPLPMRGCKDPYRVLSEVLLQOTRVEQALPYRRPFLERFPTLKALAASLE 68  
DB 32 IOWEOKRULPWRYNNDPFRIMWSEIMLQOTRVDYIDFYRMEWFPPIBELANPEE 91  
QY 69 EVLRWOGAGYRRRAEHLRLARSV-----BELPPSAELRGDLGDPYTAANAASIAFG 123  
DB 92 KLLKAMEGLGYSSRAANIQAARKQIMSEFDGEMQOTPEEISLTKGIPYTTGALASIAFG 151  
QY 124 ERVAUVCNARVRVLSRLFARSES-----PKKELPALAGLLPEGVDPVWNOALMELGATV 179  
DB 152 LPEPAVDGNVNRVSRFLCTEADIAKASSRKIFDEARKKTIIDEGHGPGEFNOAMMDLGSAL 211  
QY 180 CLPKRPRGACPLGAFPC-----RGKEAFGRYPAPRKRRAKEERLVALVLGRKGVH----- 230  
DB 212 CTPTSPKCECTCPFOAFCLANKRGQT--SPFVKTKKAKPDVYVYISALQNHSGAYYFEE 269  
QY 231 -----LERLEGRFOGLYGVPLF-----PPEELP-GREAAFGV 261  
DB 270 RDSQKLLANMWTFFPMMEVTEQEEYERLKKEMETKPEIDLFDLVAEEQNLPFEKQELFW 329  
QY 262 RSRPLGEVRHALTRRLRLEV-RCALWEGEGED-----PKRKP-----LPKLMKV 306  
DB 330 QTRHLGEVTHIFSHLKWHLVLYGRATEGAEQEFTEKNTSKMLKPADSVFEPKVMKL 389  
QY 307 L 307  
DB 390 V 390

RESULT 2  
US-09-134-000C-5115  
; Sequence 5115, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5115  
; LENGTH: 395  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-5115

Query Match 28.5%; Score 485; DB 4; Length 395;  
Best Local Similarity 33.5%; Pred. No. 2,3e-40;  
Matches 115; Conservative 56; Mismatches 108; Indels 64; Gaps 8;

QY 2 EAMRKA-----LLAWYRENARPLPMRGCKDPYRVLSEVLLQOTRVEQALPYRRFL 53  
DB 9 EAWSPAKYSSFOEDFLAWYERKKNLPRANTDAYRIMISIMLQOTRVDYIDYFYRFM 68  
QY 54 ERFPITKALAAASLEVLRWOGAGYRRRAEHLRLARSV-----BELPPSAELRGDLP 108  
DB 69 EMFPTIQLAEAPDDKLLKAMEGAGYSSRAENLKVAQAQIVSEFGKMPDTIEDIRSLKG 128  
QY 109 LGPPTAAVASTIAGERAAVDGNVRVLSRLFARSES-----PKKELPALAGLLPEGV 164  
DB 129 IGPYTAAGISIAINLEPAIDGNVRVSRFLCTEADIAKASSRKIFDEARKKTIIDRER 188  
QY 165 PGVWNOALMELGATVCLPKRPRGACPLGAFCRGKEAP--GRYPAPRRK-RAKEERLVAL 221

DB 189 PGDFNOALMDLGSAVCTPTSPKCESCPLQOYCAAOADKMTAVPVYKSKVKKPDDVYVGT 248  
QY 222 VLLGRGVNLELEGRFQG--LYGVPLPPEELPGRS----- 256  
DB 249 ILENKQOEPFL--LEQRPETGLLANMMLPPIEISIKKQFOQLKAPATEKQTLLELP 306  
QY 257 -----AAFGVSRPLGEVRHALTRRLRLEV 282  
DB 307 VTEPLVAEEPVNFPTDYETVWQKRTLGEVNHIFSHLKNHLLV 349

RESULT 3  
US-08-813-574-2  
; Sequence 2, Application US/08813574  
; Patent No. 6033473  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Ying-Fei  
; TITLE OF INVENTION: Human Muty  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/813,574  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/013,132  
; FILING DATE: 11-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Han, William T  
; REGISTRATION NUMBER: 34,344  
; REFERENCE/DOCKET NUMBER: ATG50002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-5219  
; TELEFAX: 610-270-4026  
; TEXT:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 535 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-813-574-2

Query Match 28.3%; Score 481.5; DB 3; Length 535;  
Best Local Similarity 36.5%; Pred. No. 8e-40;  
Matches 135; Conservative 46; Mismatches 106; Indels 83; Gaps 13;

QY 3 AMRKALLAWYRENARPLPMRG-----EKDPYRVLSEVLLQOTRVEQALPYRRFLE 54  
DB 81 AFRGSLLSWVDQKRLPWRRAEDMDLDRAYAVWSEVMLQOTVAVINYTGMMQ 140  
QY 55 RFPITKALAAASLEVLRWOGAGYRRRAEHLRLARS-VEEL---PPPSAEL--RGDLP 108  
DB 141 KMPTLDLASASLEEVNOIMAGIGYSSRGRRLQEGARKVVEELGGMPTTAETLQQLPG 200  
QY 109 LGPPTAAVASTIAGERAAVDGNVRVLSRLFARSES-----RESPKEELPALAGLLPEGV 163  
DB 201 VGRYTGALISIAIFGATGVVDGNVAVRVLCRVAITAADSSSTIVSQGLAQQLV-DPA 259  
QY 164 DPGVWNOALMELGATVCLPKRPRGACPLGAFCRGKE-----A 201

Db 260 RPDGPNAMMELGATVCTPQRPLCSQPVESLGRARQVBOQLASGLSSGPDVEBCA 319  
Qy 202 P-----GRYPARKRAK---BERLVALV-----LGRGVHLER- 233  
Db 320 PNTGQCHLCPSEEPDQTLGVNPPKASRKPRBESSATCUEOPGALQILVORP 379  
Qy 234 LGRFQGLYGV--LPPPEELPGRBAF-----GVRSRPLGEVNHATLHRLRV 280  
Db 380 NSGLLAGLWEPFVTEWEPSEQLORKALLQELQWAGPLPATHLHGLGEVHTSHIKLTY 439  
Qy 281 EYRGALMEGB 290  
Db 440 QVYGLALEGQ 449

RESULT 4  
US-09-252-991A-23310  
; Sequence 23310, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23310  
; LENGTH: 470  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23310

Query Match 28.1%; Score 477.5; DB 4; Length 470;  
Best Local Similarity 36.2%; Pred. No. 1,7e-39;  
Matches 126; Conservative 58; Mismatches 119; Indels 45; Gaps 14;

Qy 2 EAMRKALLAMY-RENARPLPWRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERPTLK 60  
Db 119 EGNNGAVLMDYRHRGKDLPMQGIPIYVWVSEIMLQOTQVSTVLGIDRFMAALPDVE 178  
Qy 61 AIAAASLEEVLRVWQAGYRRAEHLRLA-----RSVEELPSPFALRGPLGPGYTA 115  
Db 179 AIAAAAEDEVLRHMTGLGYSSRARNHKTQAQIVERRHAGEPRPDVQLAELPGIGSTAG 238  
Qy 116 AVASIAFGERRVAAVDGNVRVLSRLFARE-----SPK-EKELFALAGLPEGDPGVNQ 170  
Db 239 AIAASISMGIRAPLIDGNVRVLRVLAQOGYGPSEPVARALWMAARFTPH-ARVNHVYQ 297  
Qy 171 ALMELGATVCLPFRPGACPLGAFCR-----GKEAFGRYPAPRKRA-KEERLVALVLG 225  
Db 298 AMMDLGATICTSKSECLLCPVSSGRALHGREAD--YQPPKRALQOKRTLMITLAN 355  
Qy 226 RKG---VHLERLGRFQGLYGVPLPPEELPGRBAFVGRSRLGEVR-----HALTHR 276  
Db 356 RQGAILLYRPPSSGLWGLWLSLPEL--DDLGLPLAARSHSLALGERRELGLTHTFSSH 413  
Qy 277 RLRVE-----VRGA-----LWGEGBDDPKRPLPLKMEVVLAKA 310  
Db 414 QLAIEPWLVAVEGAPRAVAEGDWLWNLATPP-RIGLAAPVKKLLKRA 460

RESULT 5  
US-09-198-452A-421  
; Sequence 421, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffiths, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 421  
; LENGTH: 375  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-421

Query Match 28.0%; Score 477; DB 4; Length 375;  
Best Local Similarity 38.6%; Pred. No. 1.4e-39;  
Matches 117; Conservative 44; Mismatches 112; Indels 30; Gaps 7;

Qy 6 KALLAWYRENARPLPWRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERPTLKALAAA 65  
Db 22 EALKWFEKQKSLPFRDNPPTYSVWVSEVMLQOTRAEVIDYFNQWMBRFPTESLAAA 81  
Qy 66 SLEEVLRVWQAGYRRAEHLRLAASVE-----LPPSFALRGPLGPGYTAANAASI 120  
Db 82 KEEDVIKMEGIGYSSRAHLLLEGARWMBEPFGKIPDAISLAQIRGVGPYVHAIIAF 141  
Qy 121 AFGERVAADVGNVRVLSRLFARESPEKE-----LFALAQGLPEGVDPGVNQALMEL 175  
Db 142 AFRRRAAADVGNVRVLSRLFILETISIDESTRTWSRLAQLLPP-KSPBEVLAELIEL 200  
Qy 176 GATVCLPKRPGACPLGAFCRGKEAPGRYPAPRKRAKE---BERLVALVLL-GRGVN 230  
Db 201 GACIC-KKVPOCHRCQVQACGAMRENKQFVLPVRARKKVIFLRLVALVLYDGLVVE 259  
Qy 231 LERLEGRFQGLYGVPLPPEELPGRBAFVGRSRLGEVR-----PLGEVNHATLHRL 277  
Db 260 KRPPKEMAGLYEFPYIEVEPEBGLQDIEGFTKKMELSLSPLEFLGNLKEORHAPTNRK 319  
Qy 278 LRV 280  
Db 320 VHL 322

RESULT 6  
US-08-813-574-9  
; Sequence 9, Application US/08813574  
; Patent No. 6013473  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Ying-Fei  
; TITLE OF INVENTION: Human Muty  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/813,574  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/013,132  
; FILING DATE: 11-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Han, William T  
; REGISTRATION NUMBER: 34,344  
; REFERENCE/DOCKET NUMBER: ATG50002  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5219  
 TELEFAX: 610-270-4026  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 350 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-813-574-9

Query Match 27.7%; Score 471.5; DB 3; Length 350;  
 Best Local Similarity 35.1%; Pred. No. 4.5e-39;  
 Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAWYRENAR-PLPWGGEKDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
 DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVMLQOTQVATVIFPERFMARFPTVTDLANAP 69  
 QY 67 LEEVLRVMOAGGYRRAHRLHRLARSVEEL-----PPSFAELRGLPGIGPYTAANVASIA 121  
 DB 70 LDEVHLMTGTGCGYARARNLHKAQQVATLHGKFPETFEVVALPGVGRSTAGAILSLIS 129  
 QY 122 FGERVAADVGNVRRVLSRLFARES-PKEKE-----LFALAQGLPE-GVDPGVNQAAMEL 175  
 DB 130 LGKHPFLIDGNVRRVLCYAVSGMPGKKEVENKMSLSEQVTPAVGER--FNQAMMDL 187  
 QY 176 GATVCLPRPRGACPL--GAFCRGKEAPGRYPAPRRKRAKEERLVALVLGRKGVHL-- 231  
 DB 188 GAMICTRSKPKCSLCPQNGCIAAANNSWALYPGKKPKQTLPERTYGFLLLQHEDEVLLA 247  
 QY 232 -ERLEGRFOGLYGVPLFPPEE-----LPGREAFGVSRPLGEVRHALTHRLRV----- 280  
 DB 248 QRPSPGLMGGLYCFPPQFADEESLRQWLAQROIADNLTO-LTAFRHTFSHFHLDIVPMWL 306  
 QY 281 -----EVRGALWEGEGEDP--WKRPPLKMEKVRKALPL 313  
 DB 307 PVSSFTGCMDEGNALWYNLAQPPSVGLAAPERLLQO-LRTGAPV 350

## RESULT 7

US-09-651-656-37  
 ; Sequence 37, Application US/09651656  
 ; Patent No. 6340566  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MCCUTHEN-MALONEY, SANDRA  
 ; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY  
 ; TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE  
 ; TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,  
 ; TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES  
 ; FILE REFERENCE: IL-10689  
 ; CURRENT APPLICATION NUMBER: US/09/651,656  
 ; CURRENT FILING DATE: 2000-08-29  
 ; PRIOR APPLICATION NUMBER: 60/192,764  
 ; PRIOR FILING DATE: 2000-03-28  
 ; NUMBER OF SEQ ID NOS: 106  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 37  
 ; LENGTH: 350  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-09-651-656-37

Query Match 27.7%; Score 471.5; DB 3; Length 350;  
 Best Local Similarity 35.1%; Pred. No. 4.5e-39;  
 Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAWYRENAR-PLPWGGEKDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
 DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVMLQOTQVATVIFPERFMARFPTVTDLANAP 69  
 QY 67 LEEVLRVMOAGGYRRAHRLHRLARSVEEL-----PPSFAELRGLPGIGPYTAANVASIA 121

DB 70 LDEVHLMTGTGCGYARARNLHKAQQVATLHGKFPETFEVVALPGVGRSTAGAILSLIS 129  
 QY 122 FGERVAADVGNVRRVLSRLFARES-PKEKE-----LFALAQGLPE-GVDPGVNQAAMEL 175  
 DB 130 LGKHPFLIDGNVRRVLCYAVSGMPGKKEVENKMSLSEQVTPAVGER--FNQAMMDL 187  
 QY 176 GATVCLPRPRGACPL--GAFCRGKEAPGRYPAPRRKRAKEERLVALVLGRKGVHL-- 231  
 DB 188 GAMICTRSKPKCSLCPQNGCIAAANNSWALYPGKKPKQTLPERTYGFLLLQHEDEVLLA 247  
 QY 232 -ERLEGRFOGLYGVPLFPPEE-----LPGREAFGVSRPLGEVRHALTHRLRV----- 280  
 DB 248 QRPSPGLMGGLYCFPPQFADEESLRQWLAQROIADNLTO-LTAFRHTFSHFHLDIVPMWL 306  
 QY 281 -----EVRGALWEGEGEDP--WKRPPLKMEKVRKALPL 313  
 DB 307 PVSSFTGCMDEGNALWYNLAQPPSVGLAAPERLLQO-LRTGAPV 350

## RESULT 8

US-09-650-855-37  
 ; Sequence 37, Application US/09650855  
 ; Patent No. 6365355  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MCCUTHEN-MALONEY, SANDRA  
 ; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY  
 ; TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA  
 ; TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA  
 ; FILE REFERENCE: IL-10284  
 ; CURRENT APPLICATION NUMBER: US/09/650,855  
 ; CURRENT FILING DATE: 2000-08-29  
 ; PRIOR APPLICATION NUMBER: 60/192,764  
 ; PRIOR FILING DATE: 2000-03-28  
 ; NUMBER OF SEQ ID NOS: 106  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 37  
 ; LENGTH: 350  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-09-650-855-37

Query Match 27.7%; Score 471.5; DB 3; Length 350;  
 Best Local Similarity 35.1%; Pred. No. 4.5e-39;  
 Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAWYRENAR-PLPWGGEKDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
 DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVMLQOTQVATVIFPERFMARFPTVTDLANAP 69  
 QY 67 LEEVLRVMOAGGYRRAHRLHRLARSVEEL-----PPSFAELRGLPGIGPYTAANVASIA 121  
 DB 70 LDEVHLMTGTGCGYARARNLHKAQQVATLHGKFPETFEVVALPGVGRSTAGAILSLIS 129  
 QY 122 FGERVAADVGNVRRVLSRLFARES-PKEKE-----LFALAQGLPE-GVDPGVNQAAMEL 175  
 DB 130 LGKHPFLIDGNVRRVLCYAVSGMPGKKEVENKMSLSEQVTPAVGER--FNQAMMDL 187  
 QY 176 GATVCLPRPRGACPL--GAFCRGKEAPGRYPAPRRKRAKEERLVALVLGRKGVHL-- 231  
 DB 188 GAMICTRSKPKCSLCPQNGCIAAANNSWALYPGKKPKQTLPERTYGFLLLQHEDEVLLA 247  
 QY 232 -ERLEGRFOGLYGVPLFPPEE-----LPGREAFGVSRPLGEVRHALTHRLRV----- 280  
 DB 248 QRPSPGLMGGLYCFPPQFADEESLRQWLAQROIADNLTO-LTAFRHTFSHFHLDIVPMWL 306  
 QY 281 -----EVRGALWEGEGEDP--WKRPPLKMEKVRKALPL 313  
 DB 307 PVSSFTGCMDEGNALWYNLAQPPSVGLAAPERLLQO-LRTGAPV 350

## RESULT 9

US-09-543-681A-11959

ORGANISM: Klebsiella pneumoniae

Query Match 26.9%; Score 457.5; DB 4; Length 516;  
Best Local Similarity 34.8%; Pred. No. 2e-37;  
Matches 120; Conservative 63; Mismatches 119; Indels 43; Gaps 14;

QY 8 LLAMYREAR-PLPRGKEDPRVLVSEVLLOOTRVEQALPYRRRLERPPYLKALAAAS 66  
DB 176 VLDWDXDKGRKTLTPMQIAKTPYKWLSEVMLDQOTQVTTVPIYFERMARPPVVDLANAP 235  
QY 67 LEEVLRVQAGGYRAEHLHRLARSV-----ELPSPFAELRGLGGLGYTAAVASIA 121  
DB 236 LDEVLHMTGTGTYARARNLHKAQOAVATLHGGEFPRRTDEVAAALGVGRSTGAILSTLS 295  
QY 122 FGEVAAVDGNVRVLSLFA-----RESPKKELEFALAQGLLP-EGVDGVMQALMEL 175  
DB 296 LGQHYPIIDGNVKRYLARCVAVSGMPKKEVEKRLMDISEEVTPAQVER--FNQAMMDL 353  
QY 176 GATVCLPRPCGACPL--GAFCRGKEAGRPAPRKRAKEERLVALVLGKG--VHL 231  
DB 354 GAWCTRSKPCPCPLNSGCVAAVANHSAEYFGKKPKQLPER--TGYPFLMHGDEVFL 412  
QY 232 ERLB--GRFOGLYGVLPFPE-ELPGRBAFGVRSRPLGEV---RHALTRRRLV----- 280  
DB 413 SQRPPEVGLMGILFCPPQFADEHELEBWLKORQIKADNLVQLTRFRHTFSFHLDIVEMWL 472  
QY 281 -----EVRGALMEGEGEDP---WKRPLKMLEKVLRKALPL 313  
DB 473 TVHSSGAWMDEGNMLMNYLQAQPPSVGLAAPVRLLDQ-LKAGAPV 516

RESULT 11  
US-09-543-681A-5315  
Sequence 5315, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 5315  
LENGTH: 347  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-5315

Query Match 26.5%; Score 450; DB 4; Length 347;  
Best Local Similarity 33.2%; Pred. No. 6.4e-37;  
Matches 115; Conservative 64; Mismatches 123; Indels 44; Gaps 12;

QY 2 EAWKALLAMVREAR-PLPRGKEDPRVLVSEVLLOOTRVEQALPYRRFLERPPYLK 60  
DB 6 QQFQGVLDWVHHKQKRTKLPWQGEKTPYHVMLSVWLQOTQVATVPIYFERFIARFNV 65  
QY 61 ALAAASLEVLRVQAGYRRAEHLHRLARSV-----ELPSPFAELRGLGGLGYTAA 115  
DB 66 ALAPARLDEVLHMTGLTYARARNLHKAQHIVDHGQGFQPTDFDVCLPGVGRSTAG 125  
QY 116 AVASIAFERAAVDGNVRVLSLFARES-SKEK-----LPAALAGLLP-EGVDPEVMN 169  
DB 126 AILSLSKRPYPIIDGNVKRYLARCVAVEGMPKKEVEKNKLWEISQVPTKGE--YFN 183  
QY 170 QALMELGATVCLPRPCGACPLGAF-----RGKEAGRPAPRKRAKEERLVALVL 223  
DB 184 QAMMDLGAMVCTRIKPCKELCPLNTGCTAIVANHSAVDYGGKR--KAIPEKTYFLIL 240  
QY 224 LGRKGVHLERL--EGRFQGLYGVLPFPE---ELPGRBAFGVRSRPLGEVRLHTRR 277

```

Db      241 QYDNLVWLDKRRPAGLWGGLPAPQFETKYLLEQMLAENHOLENSKEQSLSPHTSHH 300
Qy      278 LRV-----EVRGALMEGEGEDPMKRPPLPKLMEKYLAK 309
Db      301 LDIVICVKLSTFTSLMEAOQGLWYNL-QTPASVGLAAPEVNLRLQ 345

RESULT 12
US-09-134-001C-4684
; Sequence 4684, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4684
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4684

Query Match      26.0%; Score 443; DB 3; Length 360;
Best Local Similarity 34.3%; Pred. No. 3,4e-36;
Matches 103; Conservative 56; Mismatches 119; Indels 22; Gaps 7;

Qy      3  ANRKALLMYRENAPRLPMRGEKDPYRVLVSEVLLOQTRVEQALPYRRFLERFPPTLAKL 62
      19  SRKKQIEDHFHKNQDMRRETTPYIWLSEVMLQOTQVNVIDYIRFIHREFTIGSL 78
Db      63  AAASLEEVLRVQAGAGYRRARSHLRLASVE-----ELPSPFAELRGLPGIGPTTAAV 117
Qy      79  SENNEBEVLKVMGEGLYSRANPHTAVKEVNNYDGEVYPDESFKLKGVPYTOAAV 138
Db      118  ASIAFGERVAAVDGNVRVLSRL--PARSPKEKELFLAAGLLLEGV--DPGVNQALME 174
Qy      139  MSIAFMHPLATVDGNVFRVMSRLNDYRDKLQSTRKAEQELHRYVLKDACTPQANME 198
Db      175  LGATVCLPKRPRGACPLGAFRCGEAPGRYPAPRRRAKEERLV---ALVILGRKGYHL 231
Qy      199  LGLAVCTPKSPPLCLFPCIOEHCEAFHMGTTQELPVTKSLKKTIQKVFILRNNGOYL 258
Db      232  --ERLEGRFOGLYVPLFPPEELPGREA-----AFGVRSRPLGVRNALTHRLRLAVEV 282
Qy      259  LEKRKEKELNGMWQFPM--REQTNANDVISDDLKSIETINEVFLKQFTHLWEIRV 316
Db

RESULT 13
US-09-328-352-4982
; Sequence 4982, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4982
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4982

```

```

Query Match      26.0%; Score 442; DB 4; Length 346;
Best Local Similarity 33.3%; Pred. No. 4.1e-36;
Matches 113; Conservative 60; Mismatches 118; Indels 48; Gaps 11;

Qy      7  ALLAWRENAR-PLPMRGEKDPYRVLVSEVLLOQTRVEQALPYRRFLERFPPTLAKLAA 65
      11  ALLWMPQHGHRNDPQVADDPYKVVSEIMLQOTQVKTLYQYFDEFMERFPVLEALGYA 70
Db      66  SLEEVLRVQAGAGYRRARSHLRLASVE---LPSFAELRGLPGIGPTTAAVASIAF 122
Qy      71  TWDEYAPRWAGLGYAARNLHKAAGLVAOQGFPELTIEWILPGLIGSTAGALMSLGL 130
Db      123  GERVAADVGNVRVLSRLFARES-----PKKEFLAALQGLBEGVDPGVWNOALMELGA 177
Qy      131  RQYGYIMDGNVKRVLARFALEDLSKPKQHEREMWKLAEELCTHRNHD--YTOAIMDLGA 189
Qy      178  TVCLPKRPRGACPLGAFG---RGKEAPGRYPAPRRRAKEERLVNLVILGRKGYHLR 233
Db      190  TICTEPKKELCYCPQAHCOAYOQGLEDELPFKKPKTPPVKTADVLLITQCEDEMFQOR 249
Qy      234  -LEGFQGLXGVPLFPRPE-ELPGRGAAGVRSRP-LGEVHALTH----- 275
Db      250  QAHGLMGLFCPLIENENHERLKSQPFKLOPQOTQISHFTFTMLNAHFHVEPD 309
Qy      276  --RRLREVRGALMEGEGEDPMKRPPLPKLMEKYLRLKALP 312
Db      310  QKXHLLALEEG-----QWLSP-----EQALAKGP 334
Db

RESULT 14
US-09-629-222A-36
; Sequence 36, Application US/09629222A
; Patent No. 6599700
; GENERAL INFORMATION:
; APPLICANT: Bellacosa, Alfonso
; TITLE OF INVENTION: Methods for Detection of Transition
; FILE REFERENCE: PCCC 96-21
; CURRENT APPLICATION NUMBER: US/09/629,222A
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/463,891
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US98/15828
; PRIOR FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 60/053,936
; PRIOR FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatsSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 188
; TYPE: PRT
; ORGANISM: E. coli
US-09-629-222A-36

Query Match      22.9%; Score 389.5; DB 4; Length 188;
Best Local Similarity 44.8%; Pred. No. 3.3e-31;
Matches 81; Conservative 37; Mismatches 50; Indels 13; Gaps 5;

Qy      27  PYRVLVSEVLLOQTRVEQALPYRRFLERFPPTLAKLAASLEEVLRVQAGAGYRRARSHL 86
      2  PYKWLSEVMLQOTQVATVLPYFERFVARPPTVDLANPFLDVLMLGLGLYARARNL 61
Db      87  HRLARSYBEL-----PSPFAELRGLPGIGPTTAAVASIAFGERVAAVDGNVRVLSRL 141
Qy      62  HKAQOAVATLHGKFPETFEVVALPVGVSSTGAILISLGGHFIIDGNVVRVLARCY 121
Qy      142  ARES-PKEKE---LFLAALQGLPE-GVDPGVWNOALMEIGATVCLPKRPRGACPLGAF 195
Db      122  AVSGWPGKKEVENKLMISLSEQVTDAVGVER--FQOAMMDIGAMICTRSEKSCSLCPLONG 179
Qy      196  C 196
Db      180  C 180

```



## RESULT 15

US-09-540-236-2578  
 ; Sequence 2578, Application US/09540236  
 ; Patent No. 6673910

## GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.  
 TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO MORAXELLA CATAR  
 FILE REFERENCE: 2709.2005-001  
 CURRENT APPLICATION NUMBER: US/09/540,236  
 NUMBER OF SEQ ID NOS: 3840  
 SEQ ID NO 2578  
 LENGTH: 414  
 TYPE: PRT  
 ORGANISM: M.catarhalis

US-09-540-236-2578

Query Match 20.9%; Score 355; DB 4; Length 414;

Best Local Similarity 28.5%; Pred. No. 3e-27;

Matches 109; Conservative 57; Mismatches 132; Indels 84; Gaps 13;

```

QY 3 AMRKALLAWYREAR-PLPRGKEDP---YRVLSVLLQOTRVEQALPYRRPLERP 57
DB 33 SFAKRLITFELHGRHGLPWQIHQPSADIYAVWVSEIMLOQTQVTVLKFEFFLARA 92
QY 58 TLKALAAASLEBVLRVWOGAGYVRAEHLRLASVE-----LPPSFABLRGLPGLG 110
DB 93 TVQELAVADWQEVASFVAGIGYARARNLHAGAQVADFDTHGRFPETVNEQAVKGV 152
QY 111 PYTAAVASIAGGERVAADVGRVLSRLFA----RSPKKEKLPAALQGLPEGVD 165
DB 153 RSTAGIVAMGVKKFVIGDGNVKVLAHRAVCGDITSDTKRLMEIATALTPEKYS- 211
QY 166 GWNQALMELGATVCLPKRPGACPLGAFCRGKEAPGRYPAPRKRAKE--RLVALV 222
DB 212 GHYAQMMDLGAITICRTTPKCHLCVPTDCAIYALGVQSOLPVKKKAPPKPHMSTALS 271
QY 223 LIG---RKGVHLERLEGRFQGLYGVPLFPPEELP-----GREAA 258
DB 272 LTHCGTLWLHRONGGIGWDGLMSLPF--MLPLDNDQKLDNNTLSDAIFKAMQSDKV 328
QY 259 FGVRSRPLGEV-----RHALTY-----RLRVEYRGA---- 285
DB 329 HDLHLSOLEIPIPTQTLTAYLRHTLTIVHMLYGMISCLNNSQFNQINQTLTGIDY 388
QY 286 LWESEGBDPWKRPLPKLMEKVL 307
DB 389 LMT---DTFYNLPLPAAMHKL 407

```

Search completed: January 20, 2005, 06:10:41  
 Job time : 29 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2005, 05:59:41 : Search time 78 Seconds  
(without alignments)  
1505.372 Million cell updates/sec

Title: US-09-938-901a-2

Perfect score: 1701

Sequence: 1 MEAMRKALLAWYRENNRPLP.....YARKLPLAHGVPLPDA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1697	99.8	325	10	US-09-938-901-2
2	483.5	28.4	486	9	US-09-925-301-1336
3	477	28.0	375	15	US-10-289-762-421
4	448	26.3	469	17	US-10-425-115-277336
5	448	26.3	482	15	US-10-425-114-68280
6	442	26.0	474	16	US-10-437-963-138271
7	389.5	22.9	188	15	US-10-629-951-36
8	377	22.2	313	14	US-10-156-761-12241
9	365	21.5	293	9	US-09-738-626-6433
10	337	19.8	1719	16	US-10-437-963-103808
11	263.5	15.5	184	15	US-10-424-599-248081
12	240.5	14.1	188	15	US-10-629-951-34
13	234	13.8	230	15	US-10-335-977-7283

14	228.5	13.4	187	15	US-10-335-977-7282	Sequence 7282, Ap
15	200.5	11.8	98	15	US-10-335-977-7279	Sequence 7279, Ap
16	200.5	11.8	98	15	US-10-335-977-7280	Sequence 7280, Ap
17	200.5	11.8	214	15	US-10-335-977-7461	Sequence 7461, Ap
18	199	11.7	268	9	US-09-864-866-43	Sequence 43, Appl
19	191	11.2	185	15	US-10-629-951-35	Sequence 35, Appl
20	184.5	10.8	310	14	US-10-156-761-13127	Sequence 13127, A
21	176.5	10.4	90	11	US-09-864-408A-9042	Sequence 9042, Ap
22	176	10.3	259	14	US-10-369-493-5691	Sequence 5691, Ap
23	163.5	9.6	304	14	US-10-424-599-280749	Sequence 280749, Ap
24	158.5	9.3	355	14	US-10-369-493-2113	Sequence 2113, Ap
25	155	9.1	172	15	US-10-424-599-150617	Sequence 150617, A
26	155	9.1	193	16	US-10-437-963-113667	Sequence 113667, A
27	150	8.8	373	15	US-10-425-114-65596	Sequence 65596, A
28	149	8.8	366	17	US-10-425-115-305120	Sequence 305120, A
29	147.5	8.7	260	9	US-09-738-626-3828	Sequence 3828, Ap
30	145	8.5	211	9	US-09-912-020-296	Sequence 296, App
31	145	8.5	211	17	US-10-771-241-296	Sequence 296, App
32	137.5	8.1	184	15	US-10-629-951-33	Sequence 33, Appl
33	130	7.6	158	15	US-10-424-599-246739	Sequence 246739, A
34	129.5	7.6	170	16	US-10-767-701-38640	Sequence 38640, A
35	129.5	7.6	5245	14	US-10-328-079-11	Sequence 11, Appl
36	128	7.5	281	14	US-10-128-714-3128	Sequence 3128, Ap
37	128	7.5	461	14	US-10-128-714-8128	Sequence 8128, Ap
38	126	7.4	272	14	US-10-369-493-13299	Sequence 13299, A
39	125.5	7.4	148	16	US-10-767-701-61161	Sequence 61161, A
40	125.5	7.4	6238	15	US-10-343-710-71	Sequence 71, Appl
41	120.5	7.1	735	17	US-10-425-115-216806	Sequence 216806, A
42	120.5	7.1	776	15	US-10-425-114-70995	Sequence 70995, A
43	119	7.0	187	15	US-10-425-114-68053	Sequence 68053, A
44	118.5	7.0	416	15	US-10-425-114-43512	Sequence 43512, A
45	117.5	6.9	1072	16	US-10-437-963-114893	Sequence 114893, A

#### ALIGNMENTS

RESULT 1	US-09-938-901-2	Application US/09938901
1	Sequence 2, Appl	
2	Publication No.	US20030008291A1
3	GENERAL INFORMATION:	
4	APPLICANT:	Kuramitsu Seiki,
5	TITLE OF INVENTION:	GENE ENCODING DNA REPAIR ENZYME
6	FILE REFERENCE:	PH-1261-US
7	CURRENT APPLICATION NUMBER:	US/09/938,901
8	CURRENT FILING DATE:	2001-08-24
9	PRIOR APPLICATION NUMBER:	JP2001-47762
10	PRIOR FILING DATE:	2001-02-23
11	NUMBER OF SEQ ID NOS:	17
12	SOFTWARE:	Patentin Ver. 2.0
13	SEQ ID NO 2	
14	LENGTH:	325
15	TYPE:	PRT
16	ORGANISM:	Thermus thermophilus
17	US-09-938-901-2	
18	Query Match	99.8%; Score 1697; DB 10; Length 325;
19	Best local Similarity	99.7%; Pred. No. 1.8e-145;
20	Matches 324;	Conservative 1; Mismatches 0; Indels 0; Gaps 0;
21	QY	1 MEAMRKALLAWYRENNRPLPMPGEXKDPYRVLSVLLQOTRVBQALPYRRFLERFPTLK 60
22	DB	1 VAMRKALLAWYRENNRPLPMPGEXKDPYRVLSVLLQOTRVBQALPYRRFLERFPTLK 60
23	QY	61 AAAAALEEVLEWQAGYRRRAEHRLARVVEEIPFAFRLRGIPGCPYTAAVASI 120
24	DB	61 AAAAALEEVLEWQAGYRRRAEHRLARVVEEIPFAFRLRGIPGCPYTAAVASI 120
25	QY	121 AFGERVAADVGVRRVRLSLFARSPKEKELPALAQGLPEGVDPGVNQAALMELGATVC 180
26	DB	121 AFGERVAADVGVRRVRLSLFARSPKEKELPALAQGLPEGVDPGVNQAALMELGATVC 180

```

QY      181  LPRKPRGACPLGAFGRKGAAPGRYAPARKRAKEERLVALVILGRKGVHLERLEGROG 240
      181  LPRKPRGACPLGAFGRKGAAPGRYAPARKRAKEERLVALVILGRKGVHLERLEGROG 240
Db      241  LYGVLFPPEELPGREAAFGVRSRPLGEVRLHTRRLAVRGAALMEGEGDPMKRPLP 300
QY      241  LYGVLFPPEELPGREAAFGVRSRPLGEVRLHTRRLAVRGAALMEGEGDPMKRPLP 300
Db      241  LYGVLFPPEELPGREAAFGVRSRPLGEVRLHTRRLAVRGAALMEGEGDPMKRPLP 300
QY      301  KLMEKVLKALPLLAHAGVPLPDA 325
      301  KLMEKVLKALPLLAHAGVPLPDA 325
Db      301  KLMEKVLKALPLLAHAGVPLPDA 325

RESULT 2
US-09-925-301-1326
/ Sequence 1326, Application US/09925301
/ Patent No. US2002052308A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
/ FILE REFERENCE: PA106
/ CURRENT APPLICATION NUMBER: US/09/925,301
/ PRIOR FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: PCT/US00/05862
/ PRIOR FILING DATE: 2000-03-08
/ PRIOR APPLICATION NUMBER: 60/124,270
/ PRIOR FILING DATE: 1999-03-12
/ NUMBER OF SEQ ID NOS: 1694
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 1326
/ LENGTH: 486
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (34)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (438)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (447)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1326

Query Match      28.4%; Score 483.5; DB 9; Length 486;
Best Local Similarity 36.5%; Pred. No. 4,5e-35;
Matches 135; Conservative 46; Mismatches 106; Indels 83; Gaps 13;

QY      3  AMRKALAMYRENNARPLPMRG-----EKDPYRVLYSEVLLQOTRVEQALPYRRFLP 54
      3  AMRKALAMYRENNARPLPMRG-----EKDPYRVLYSEVLLQOTRVEQALPYRRFLP 54
Db      106  AFRGSLSMYDQERKDLFWRRRAEDMDLDRRAVAWVSEVLLQOTRVAIVINYTGWMG 165
QY      55  RPPFLKALAAASLEEVNMQAGYVRAEHLRLARS-VEEL--PSPFAEL--RGLRG 108
      55  RPPFLKALAAASLEEVNMQAGYVRAEHLRLARS-VEEL--PSPFAEL--RGLRG 108
Db      166  KMPFLQDIALASLEEVNMQAGYVRAEHLRLARS-VEEL--PSPFAEL--RGLRG 108
QY      109  LGPYTAAVASTAFGEVAAVDGNVRYLSRLFA-----RSPREKELPLAQLPPEGV 163
      109  LGPYTAAVASTAFGEVAAVDGNVRYLSRLFA-----RSPREKELPLAQLPPEGV 163
Db      226  VERTYTAGATASIAFGAGTGVVDGNVRYLSRLFA-----RSPREKELPLAQLPPEGV 163
QY      164  DRGVNQAALMELGATVCLPKPRGACPLGAFGRKGE-----A 201
      164  DRGVNQAALMELGATVCLPKPRGACPLGAFGRKGE-----A 201
Db      285  RPDDEFQAALMELGATVCLPKPRGACPLGAFGRKGE-----A 201
QY      202  P-----GRYPARKRAK-----EELVALYL-----LGRKGVHLER- 233
      202  P-----GRYPARKRAK-----EELVALYL-----LGRKGVHLER- 233
Db      345  PNTGQCHLCPREEPMDOTLGVNFPKRSRKRPRESSATVLEQPALGAQILVGRP 404
QY      234  LGRFPGQLGVPR--LPPPEELPGREAAF-----GVNSRPLGEVRLHTRRLRV 280
      234  LGRFPGQLGVPR--LPPPEELPGREAAF-----GVNSRPLGEVRLHTRRLRV 280
Db      405  NSGLLAGLWEPFVTPWESQLOKRLLOLQKXAGPLPATIXRHILGEVHTFSHIKLT 464

```

```

QY      281  EYRGALMEGE 290
      281  EYRGALMEGE 290
Db      465  QYVGLALBEGQ 474

RESULT 3
US-10-289-762-421
/ Sequence 421, Application US/10289762
/ Publication No. US20040006218A1
/ GENERAL INFORMATION:
/ APPLICANT: Griffiths, R.
/ TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
/ FILE REFERENCE: 9710-003-999
/ CURRENT APPLICATION NUMBER: US/10/289,762
/ PRIOR FILING DATE: 2003-03-27
/ NUMBER OF SEQ ID NOS: 6849
/ SEQ ID NO 421
/ LENGTH: 375
/ TYPE: PRT
/ ORGANISM: Chlamydia pneumoniae
US-10-289-762-421

Query Match      28.0%; Score 477; DB 15; Length 375;
Best Local Similarity 38.6%; Pred. No. 1,3e-34;
Matches 117; Conservative 44; Mismatches 112; Indels 30; Gaps 7;

QY      6  KALLAMYRENNARPLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAA 65
      6  KALLAMYRENNARPLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAA 65
Db      22  EALKKMFERNKRSPLPMRNDPPIYVWSEVLLQOTRVEQALPYRRFLERFPTLKALAA 81
QY      66  SLEEVLYRWQAGYVRAEHLRLARSVE-----LPPFAELRGLPGAGPYTAAVAST 120
      66  SLEEVLYRWQAGYVRAEHLRLARSVE-----LPPFAELRGLPGAGPYTAAVAST 120
Db      82  KEEDVYKLMESGLGYVSRARHLLEGARWMEBFHGKIDDAISLAQIRGVGPYTHALIA 141
QY      121  AFGERYAAVDGNVRYLSRLPARSPKEKE-----LPLAQLPPEGVNDGNVQAALMEL 175
      121  AFGERYAAVDGNVRYLSRLPARSPKEKE-----LPLAQLPPEGVNDGNVQAALMEL 175
Db      142  AFRRAAAVDGNVRYLSRLPARSPKEKE-----LPLAQLPPEGVNDGNVQAALMEL 175
QY      176  GATVCLPKPRGACPLGAFGRKGAAPGRYAPARKRAKE-----EELVALYL--GRKGV 230
      176  GATVCLPKPRGACPLGAFGRKGAAPGRYAPARKRAKE-----EELVALYL--GRKGV 230
Db      201  GATCIC-KYCPQCHRCVRCACGAMRENKQFVLPVHARKVIFLHRLVALVLYGSLVE 259
QY      231  LERLEGRFQGLGVPLFPPEELPGREAAFGVRSR-----PLGEVRLHTRRLRV 277
      231  LERLEGRFQGLGVPLFPPEELPGREAAFGVRSR-----PLGEVRLHTRRLRV 277
Db      260  KRRPKEMMAGLYEFPYIEVEPEBGLDIEGFTKGMELSLSEPLERGLVKEQRHAFTHWK 319
QY      278  LRV 280
      278  LRV 280
Db      320  VHL 322

RESULT 4
US-10-425-115-277336
/ Sequence 277336, Application US/10425115
/ Publication No. US20040214272A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53222)B
/ CURRENT APPLICATION NUMBER: US/10/425,115
/ PRIOR FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 369326
/ SEQ ID NO 277336
/ LENGTH: 469
/ TYPE: PRT
/ ORGANISM: Zea mays

```

```
FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_184509C.1.pep
; US-10-425-115-277336

Query Match      26.3%; Score 448; DB 17; Length 469;
Best Local Similarity 36.4%; Pred. No. 7.2e-32;
Matches 123; Conservative 44; Mismatches 113; Indels 58; Gaps 10;

QY 3 AMRKALLANREARPLPWR-----GKDPYRLVSEVLLQOTRVEQALPYRRFLERFP 57
DB 57 ALRAQQLRWYDARRDLPMRCVSGSEEBRAYAVWSEVMLQOTRVPVVAYYRRMARWP 116
QY 58 TLKALAAASLEEVLRVWOGAGYRRARHRLARSVEE---LPPSFALRGFLGCLPYTA 114
DB 117 TVRSLLAAATQOEVENEMAGLGYRRARFLLEGAKQIIEKLFECTALALREVGIGIDYTA 176
QY 115 AAVASIAFGERVAADVGNVRRVLSRLFA--RESPKEKELFALAGLLPEGVD---PGVMNQ 170
DB 177 GAIASIAFNEBVVVDGNVIRVLSRLYTTADNPKESSTVVRFPMDLVGQWVDPLRPGDFNQ 236
QY 171 ALMELGATVCLPKRPGCGACPLGAFG-----RGKAPGRYPAPR---KRRAKEERLVAL 221
DB 237 AMMELGATLCSKTKPGCSQCPVSSHCQALALSREKSSVQVTDPPRVVPKAKPRSDPAAYC 296
QY 222 VILGRKGV-----HLERL-----EGRFGLYGVPLFPPE---ELPGRBAFG- 260
DB 297 VVOIAGGLEBEAADPKGNHLLFLIKRPEGLLAGLMEFPVLVVDGKTDLLNRRKAMDK 356
QY 261 -----VRSRPLGEVRLHTRRLRYEV 282
DB 357 YLSKLISIDMVRKPDVILREDVGEHVFHSIRLTMHV 394

RESULT 5
US-10-425-114-68280
; Sequence 68280, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaka, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68280
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB7326SD07_FLI.pep
; US-10-425-114-68280

Query Match      26.3%; Score 448; DB 15; Length 482;
Best Local Similarity 36.4%; Pred. No. 7.5e-32;
Matches 123; Conservative 44; Mismatches 113; Indels 58; Gaps 10;

QY 3 AMRKALLANREARPLPWR-----GKDPYRLVSEVLLQOTRVEQALPYRRFLERFP 57
DB 70 ALRAQQLRWYDARRDLPMRCVSGSEEBRAYAVWSEVMLQOTRVPVVAYYRRMARWP 129
QY 58 TLKALAAASLEEVLRVWOGAGYRRARHRLARSVEE---LPPSFALRGFLGCLPYTA 114
DB 130 TVRSLLAAATQOEVENEMAGLGYRRARFLLEGAKQIIEKLFECTALALREVGIGIDYTA 189
QY 115 AAVASIAFGERVAADVGNVRRVLSRLFA--RESPKEKELFALAGLLPEGVD---PGVMNQ 170
DB 190 GAIASIAFNEBVVVDGNVIRVLSRLYTTADNPKESSTVVRFPMDLVGQWVDPLRPGDFNQ 249
```

```
QY 171 ALMELGATVCLPKRPGCGACPLGAFG-----RGKAPGRYPAPR---KRRAKEERLVAL 221
DB 250 AMMELGATLCSKTKPGCSQCPVSSHCQALALSREKSSVQVTDPPRVVPKAKPRSDPAAYC 309
QY 222 VILGRKGV-----HLERL-----EGRFGLYGVPLFPPE---ELPGRBAFG- 260
DB 310 VVOIAGGLEBEAADPKGNHLLFLIKRPEGLLAGLMEFPVLVVDGKTDLLNRRKAMDK 369
QY 261 -----VRSRPLGEVRLHTRRLRYEV 282
DB 370 YLSKLISIDMVRKPDVILREDVGEHVFHSIRLTMHV 407

RESULT 6
US-10-437-963-138271
; Sequence 138271, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 138271
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_39677C.1.pep
; US-10-437-963-138271

Query Match      26.0%; Score 442; DB 16; Length 474;
Best Local Similarity 35.0%; Pred. No. 2.6e-31;
Matches 122; Conservative 47; Mismatches 110; Indels 70; Gaps 12;

QY 3 AMRKALLANREARPLPWRGKDP-----YRLVSEVLLQOTRVEQALPY 48
DB 56 AVRABELLRVYDANRRDLPMRRAAEPAGSGRGEGRAYAVWSEVMLQOTRVPVVDY 115
QY 49 YRRFLERFPTLKALAAASLEEVLRVWOGAGYRRARHRLARSV---BELPSPFALRG 105
DB 116 YSRMARWPTVDSLLAATQOEVENEMAGLGYRRARFLLEGAKQIIEKSEFPCTASTLRE 175
QY 106 LRLGIPYTAASIAFGERVAADVGNVRRVLSRLFA--RESPKEKELFALAGLLPEGVD 164
DB 176 VRSIGDYTAGALASIAFNEBVVVDGNVIRVLSRLYTTADNPKESSTVVRFPMDLVG 225
QY 165 ---PGVMNQALMELGATVCLPKRPGCGACPLGAFGCGKAPGR-----YP--APRKR 211
DB 236 PSRPGDFNQMAMELGATLCSKTKRPGCSQCPVSSHCQALALSQNASVAKYTDPRVVPKAK 295
QY 212 RAKEERLVALVILGRKGVHLERLEGRFQGLYGVPLFPPEBLPGRBAFGVRSRPLGEVRH 271
DB 296 PRSDPAAYCVVOIS-----QGRGEGI-----AEAEKDNVLLFLIKRP 332
QY 272 ALTHRLRVEVVGALWE-----GEGE--DPWKRPPLKMEKVLKRALPL 313
DB 333 -----BEGLLAGLMEFPVSVNBGKTDTLNR--RKMDKTIKQULSI 372

RESULT 7
US-10-629-951-36
; Sequence 36, Application US/10629951
```

Query Match	Score	DB	Length
22.2%	377	14	313

[illegible]



Db 61 QLKELARVINDYGVRRNRKAILDLPGVKYTCAAVMCLAFGKAAWVANFVRVYNR 120  
QY 140 LFARE----SPKREKELFALQGLPEGVDPGVNMOALMELGATVCLPRPRCGACPLGAF 195  
Db 121 YFGSGYENLNNHMLMELMELATLVPGGKCRD-FNLGLMDFSAIICAPRKCCEKCGSKL 179  
QY 196 C 196  
Db 180 C 180

RESULT 13  
US-10-335-977-7283  
; Sequence 7283, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: Windows NT 4.0  
; SOFTWARE: UNIX  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/335,977  
; FILING DATE: 30-Dec-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/993,002  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragoras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: GTN-018  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 7283:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 230 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...230  
; SEQUENCE DESCRIPTION: SEQ ID NO: 7283:  
US-10-335-977-7283

Query Match 13.8%; Score 234; DB 15; Length 230;  
Best Local Similarity 30.6%; Pred. No. 7,7e-13;  
Matches 66; Conservative 33; Mismatches 54; Indels 64; Gaps 9;  
QY 95 ELPPSFALRGLPGLGPTTAANAASIAFGERRVAADGNVRRVLSRLFARESPEKELFAL 154  
Db 7 QLPNDVQSLTKLKGIGAYTANAILCFGFRKSAQVDANVRVLRFLF----- 53  
QY 155 AAGLLPEGVDPGV-----WNOALMELGATVCLPRPRCGACPLGA 194  
Db 54 -----GLDPNTHAKDLQIKANDFLNLSFNHNOALIDLGLICSPK-PKCAICPFNP 105

QY 195 FCRGKEAPGRYPAPRRR-AKEERLVALVLLGRGVHLERLEGRFOGLY-GVPLFP----- 248  
Db 106 YCIGKHLNHLHTLKKQEIIGERYIGVYQNNQ-IALEKIE---QKTVIGNHFPNLKE 161  
QY 249 --PEELGREAPGVRSRPLGEVRRHALTHRRLEVEV 282  
Db 162 NLEFKLPF-----LGTIKSHHTKFKNLNL 186

RESULT 14  
US-10-335-977-7282  
; Sequence 7282, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: Windows NT 4.0  
; SOFTWARE: UNIX  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/335,977  
; FILING DATE: 30-Dec-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/993,002  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragoras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: GTN-018  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 7282:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 187 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...187  
; SEQUENCE DESCRIPTION: SEQ ID NO: 7282:  
US-10-335-977-7282

Query Match 13.4%; Score 228.5; DB 15; Length 187;  
Best Local Similarity 33.5%; Pred. No. 1.9e-12;  
Matches 59; Conservative 25; Mismatches 45; Indels 47; Gaps 7;  
QY 95 ELPPSFALRGLPGLGPTTAANAASIAFGERRVAADGNVRRVLSRLFARESPEKELFAL 154  
Db 16 QLPNDVQSLTKLKGIGAYTANAILCFGFRKSAQVDANVRVLRFLF----- 62  
QY 155 AAGLLPEGVDPGV-----WNOALMELGATVCLPRPRCGACPLGA 194  
Db 63 -----GLDPNTHAKDLQIKANDFLNLSFNHNOALIDLGLICSPK-PKCAICPFNP 114  
QY 195 FCRGKEAPGRYPAPRRR-AKEERLVALVLLGRKGVHLERLEGRFOGLY-GVPLFP 248



DB 115 YCIGKHLERHTLKKQEIQEBRYLGVV1QNNQ-IALEKIE---OKLYLGMHFP 166

## RESULT 15

US-10-335-977-7279

; Sequence 7279, Application US/10335977  
; Publication No. US20040052799A1

## GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR

DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 7279:

SEQUENCE CHARACTERISTICS:

LENGTH: 98 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...98

SEQUENCE DESCRIPTION: SEQ ID NO: 7279:

US-10-335-977-7279

Query Match 11.8%; Score 200.5; DB 15; Length 98;

Best Local Similarity 48.4%; Pred. No.2.8e-10; Mismatches 29; Indels 5; Gaps 3;

Matches 45; Conservative 14; Mismatches 29; Indels 5; Gaps 3;

QY 1 MEAMRRALLAWYRENR--PLPWR--GKDPYRVLVSEVLLQOTRYEOALP--YRRFLER 55

DB 1 LETLNHALIKWYEFGRKDLFPRLNKGINAPYEVYISEVWSQOTQISTYIERFYPPPLKA 60

QY 56 PPTLKALAAASLEVRVWQAGAYYRAEHLR 88

DB 61 PPTLKDLANAPLEEVLLWRLGVLGYSAKRLKK 93

QY 56 PPTLKALAAASLEVRVWQAGAYYRAEHLR 88

DB 61 PPTLKDLANAPLEEVLLWRLGVLGYSAKRLKK 93

QY 56 PPTLKALAAASLEVRVWQAGAYYRAEHLR 88

DB 61 PPTLKDLANAPLEEVLLWRLGVLGYSAKRLKK 93

Search completed: January 20, 2005, 06:16:04

Job time : 82 secs

**This Page Blank (uspto)**